# Chapter 4
# Tools

**Abstract**  If theoretical developments in evidence assessment are to prove useful, guidance on implementation is essential, and this chapter fills that need. A variety of tools are offered, which can be used either in isolation, or in the various combinations suggested. The starting point is an *Is your policy really evidence-based?* tool which should be very widely usable to give a very quick overview. Then two tools are offered for guideline developers for medical practice; these offer improved assessment of evidence of mechanism in assessing clinical trials, and, if needed, in basic science papers. For politicians, journalists, academics, and so on, a critical appraisal tool is offered alongside GRADE-style tables for mechanism assessment. A final tool is designed specifically for public health and social care.

In this chapter, we present a number of tools for evaluating evidence of mechanisms that have been tailored for different users. A flowchart that shows how these tools can be used together is presented below in Fig. 4.1.

## 4.1   Introduction

**How to use these tools**
For most users, the *Is your policy really evidence-based?* tool (Sect. 4.2) will be the best place to start, because it can give a quick indication of cases where a more detailed review of evidence might be valuable. If a policy is found to have possible weaknesses in its underlying evidence base, the user can then employ the other tools provided here to produce a more thorough account of the strengths and weaknesses of the policy's evidence base. While we encourage interested users to experiment with each of these tools to see which might best fit their purposes, we propose the following provisional plan:

**For those interested in guidelines for medical practice**. We would encourage these users to move on to a more systematic review of evidence using the *Mechanisms in Clinical Research appraisal tool* (see Sect. 4.3). This might also involve a more
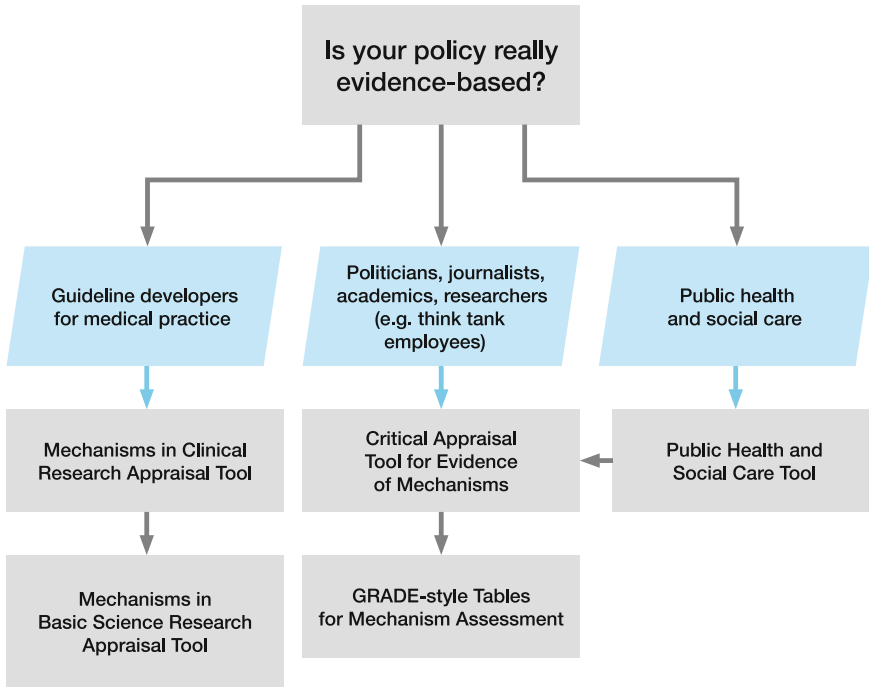
**Fig. 4.1** A suggested work-flow for using the tools presented here

detailed review of evidence arising from basic science work using the *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4).

**For those working on public health and social care guidelines**. The *Public Health and Social Care* tool (in Sect. 4.7) would be the most natural place to begin, because it explicitly asks appraisers to evaluate evidence of mechanisms that pertains both to individuals and to groups. Because of the diversity of the underlying research in public health and social care policies, the *Critical Appraisal Tool for Evidence of Mechanisms* would be the most useful tool to apply (see Sect. 4.5).

**For those interested in other policies, such as politicians, journalists, and academics**. The most natural way to proceed would depend on the nature of the policy in question. If the policy is largely medical (i.e. dealing with the effects of an intervention on an individual, with a largely biological theme) then the *Mechanisms in Clinical Research appraisal tool* would be appropriate (see Sect. 4.3), perhaps followed by the *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4). Otherwise, the *Critical Appraisal Tool for Evidence of Mechanisms* (Sect. 4.5) could be used in combination with the *GRADE-style Tables for Mechanism Assessment* (Sect. 4.6) as a next step.

**Limitations of these tools**

These tools are fallible, and their use is not a substitute for expert appraisal of a guideline or policy. Answering each of the steps requires user judgement, and the scores produced by each tool contribute to—rather than determine—the quality of recommendations. In other words, the tool alone will not provide a final and complete judgement of the quality of evidence, and their use is not a substitute for expert judgement.

These tools are specifically designed to assist in the evaluation of causal relationships. Guidance that relies on the precautionary principle may therefore score poorly, just because the precautionary principle is used when evidence of causal relationships is limited. Those poor scores should not therefore be interpreted as sufficient to alter such guidance.

These tools are currently beta versions that are suitable for testing. They have been tested by the EBM+ team during development. We welcome feedback on these tools via the EBM+ website at ebmplus.org. Feedback will help inform the next version of these tools, which will be accessible from the EBM+ website.

## 4.2 Is Your Policy Really Evidence-Based?

**Introduction**

This is a tool for appraising a wide range of policy decisions. Policies are likely to be more effective when they are based on evidence. But there are many kinds of evidence, and many ways to use evidence. Just as not all kinds of evidence are created equal, not all ways of using evidence are equally good. This tool permits the user to draw rapid but useful conclusions about the evidence that a particular policy is based on, and the way that it is based on this evidence.

Policies that use different kinds of evidence together, in an explicit and careful way, are generally better justified than policies that do not. This tool allows the user to quickly and fairly judge whether their policy is evidence-based in this way. Whilst the effectiveness of a policy is somewhat dependent on the strength of its evidence, other factors are also significant. These include proper implementation, strict adherence, and the responsiveness of policy updates.

**Who should use this tool**

This tool is a light-touch and rapid means of appraising the way that a recommendation is supported by its evidence. It is intended for use on existing policies, rather than being a tool for those constructing recommendations in the first instance. The tool was written largely with medicine and social care in mind. For example, it asks questions about evidence from basic science research because this plays an important role in supporting policy in those areas. However, we acknowledge that other types

of information are used in concert with evidence from scientific research in building policies—this tool can accommodate a wide range of different needs and different stakeholder groups working with issues of medical policy. It is envisaged that civil servants, activists, political parties, What Works Centres, and guideline developers will find this tool useful. The tool in table 4.1 might also be valuable in other areas (such as evaluating economic policy) with appropriate translation.

To provide some examples of ways that this tool might be used:

- Clinicians in primary care, public health, or social care, might use this tool as a check when considering the implementation of a new clinical guideline, or in other situations where rapid appraisals of guidelines might be otherwise helpful (for example, in multidisciplinary team meetings).
- Patient groups might use this tool to aid discussions of new treatment recommendations.
- Journalists might use the tool to begin investigating controversial policy decisions.
- Guideline authors might use this tool as a first step when considering revisions to existing guidance.
- Decision makers in local authorities might choose to use this tool when making decisions about service provision.
- Politicians (and their teams) could use this tool to evaluate their manifesto claims—or those of their opponents.
- Directors of social care and public health might use this tool to evaluate existing practices.
- This tool would be useful as part of a post-hoc effectiveness evaluation tool-kit that could be applied to policies in the event of their failure.

It is important to remember that policy evolves and develops out of many actions and involves many actors. This is true in democratic societies (where these interactions are usually at least partially visible). It is also true in more closed societies, where it is less easy to observe. In both cases, evidence and its appraisal are but one part of the mix. The relationships have been studied by political scientists (Kingdon and Thurber 1984), by policy makers themselves (National Audit Office 2003) as well as social scientists more generally (Nutley et al. 2000). There are plentiful models describing the process (Cooksey 2006; Ogilvie et al. 2009). The relationship between evidence and policy is a complex one. This has to be acknowledged, but that notwithstanding, it is important to apply the highest standards of evaluation we can to the available evidence.

**How to use this tool**
This tool should be used when examining a specific policy or recommendation. For example, we might be interested in examining a claim that, for disease $x$ in population $y$ use drug $z$. This policy will (hopefully) be supported by some group of research evidence that shows that drug $z$ is the most effective treatment for disease $x$.

**Table 4.1** Is your policy really evidence-based tool

### Question 1: Is the policy or recommendation explicit?

A recommendation can't be evidence-based unless it is explicit. For instance, the recommendation 'be more healthy' cannot be evidence-based unless it specifically describes how you should be healthy.

| | |
|---|---|
| Yes | ? |
| Somewhat | ? |
| No | ? |

### Question 2: Is the recommendation based on empirical studies of some kind?

Evidence is empirical. Unless a recommendation is based on empirical observations it cannot be evidence based. Simulations, models, and well-verified theories count as empirical evidence here.

| | |
|---|---|
| Yes (named or described studies) | ? |
| Somewhat (studies not named or described) | ? |
| No | ? |

### Question 3: Is the policy or recommendation made for a suitable population?

Different populations may have different needs. A recommendation is supported by evidence if it was produced in a group of research participants known to be similar to the target population.

| | |
|---|---|
| Yes (similar populations) | ? |
| Maybe (populations not described or known) | ? |
| No (populations known to be dissimilar) | ? |

### Question 4a: Is the recommendation based on suitable evidence of correlations?

Suitable evidence here might include clinical trials or observational studies.

| | |
|---|---|
| Yes | ? |
| Unclear or limited | ? |
| No | ? |

### Question 4b: Is the recommendation based on suitable evidence of mechanisms?

Are laboratory studies or basic science research used as part of the evidence-base for the recommendation?

| | |
|---|---|
| Yes | ? |
| Unclear or limited | ? |
| No | ? |

### Question 5: Are the sources of evidence integrated together in the recommendation?

Integrating sources of evidence leads to stronger recommendations. For example, has the selection of biomarkers in a trial been integrated with current best laboratory evidence?

| | |
|---|---|
| Yes | ? |
| Unclear or limited | ? |
| No | ? |

### Question 6: Are the limitations of the evidence explicit in the recommendation?

The explicit expression of the limitations of the evidence is usually linked to more consistent recommendations (eg. noting patterns of co-morbidity in trial subjects in guidelines).

| | |
|---|---|
| Yes | ? |
| Unclear or partial expression of limitations | ? |
| No | ? |

### Question 7: Does the policy or recommendation offer comparisons to other recommendations?

New recommendations are either intended to replace older recommendations or give guidance that is more appropriate for a particular population (geographical, demographic...). Is the recommendation compared to others in either of these ways?

| | |
|---|---|
| Yes (or not applicable) | ? |
| Unclear or partial discussion | ? |
| No | ? |

### Overall Score

- A red box in Questions 1 or 2 suggests that the recommendation is poorly evidence-based.
- A red box (or 3 yellow boxes) by Questions 3, 4 or 5 suggests that the recommendation is somewhat evidence-based.
- No red boxes (or less than 3 yellow boxes) by Questions 6 or 7 suggest that the recommendation is strongly evidence-based.

The tool then asks users a series of questions that reveal difficulties in the evidential support for that policy. These are ranked because failures in the early questions reveal more serious difficulties than failures in the later questions. These steps correspond to aspects of the account of how to gather, evaluate, and use, evidence of mechanisms that is developed in Part III. There are seven steps, each with a simple traffic-light checklist (green, yellow, or red) and each of which reflects one aspect of the relationship between the recommendation and the evidence base. The overall score for a particular policy can then be expressed by recording the lowest numbered step in which the red box is checked. For example, a policy would score 3 if it were found to be based on research on a population that was extremely unlike the intended population for its use. Note that if no red boxes are checked for any of the questions then the overall score should be noted as 7+, indicating that a policy is as evidence-based as possible. Multiple yellow flags should indicate caution, and we suggest that when three or more yellow flags are present, the score should be recorded as equal to the stage at which the third yellow flag is indicated. This overall score gives an extremely concise measure of the strength of the links between the evidence-base and the recommendation. A fuller appraisal of the policy can also be easily seen by consulting the full page of scores for each step. These initial appraisals can then form a basis for more detailed appraisal using other tools, as detailed in Sect. 4.1.

## 4.3   Mechanisms in Clinical Research Appraisal Tool

**Introduction**
This tool presents a method that a researcher would use to evaluate a group of clinical research publications. The aim of this method is to facilitate the construction of concise summaries of the mechanistic aspects of a group of clinical research publications. These summaries can then be used by a panel of experts in the context of making policy decisions about healthcare in combination with other data extraction tools (such as GRADE). Note that this tool is not intended to produce a full reconstruction of all the mechanisms that might be relevant. Instead, the summaries are intended to reveal the mechanistic aspects of clinical research. For example, some understanding of the hypothesised mechanism of action of a drug will inform the design of a clinical trial testing that drug. These mechanistic assumptions should be considered when interpreting this clinical trial.

This tool is comparatively simple, and therefore is intended for use in circumstances where the details of a mechanism are thought likely to be straightforward. In cases where either a) the consequences of a policy decision are rather serious (such as making decisions about medicines for use in pregnancy) or b) when the research base that grounds a body of clinical research is disputed or complex (such as the evaluation of treatments for chronic fatigue syndrome) we suggest that a more detailed appraisal be conducted using our *Mechanisms in Basic Science Research appraisal tool* (see Sect. 4.4). A more theoretical approach to the appraisal process can be found in Fig. 5.1.

**Who should use this tool**

This tool is intended for use during the development of clinical guidelines. Parts of this tool can be used by different groups as the process of guideline development proceeds. The data extraction parts of this tool (**A1**–**A3**, as well as **A5** if used) may be of use to literature review specialists alongside existing appraisal work. While these parts of the tool do assume some expertise in dealing with the medical literature, we do not assume domain-specific expertise in these parts. Parts of this tool—particularly **A4**—do assume a higher level of expertise in some specified scientific domain, and this stage will largely be carried out by domain experts. Finally, **A6** is intended to be carried out by those with expertise in producing guidance from clinical research.

This tool has been designed with the current (2018) practices of NICE as an archetype. We understand that practices vary in different contexts, and that the demands of a different context of practice might produce difficulties in using this tool.

**How to use this tool**

We describe a six-stage method for using this tool. The numbers of the stages (e.g. **A3**) are also shown on the flowchart in Fig. 4.2 to assist in understanding the overall appraisal process. Each of the stages will help evaluate the evidence-base that supports (or undermines) a drug's safety and efficacy. Note that not all steps will be necessary in each case. Instead, this process is adaptable to suit cases where the evidence base is favourable, or cases where the evidence base is unfavourable, or cases where the evidence base is more mixed. Note too that different stages of the process are likely to be carried out by different evaluators. We have designed this tool to assist smooth transitions between evaluators. An overview of the intended process is below.

**A1**: **collate clinical studies** At this stage, the process is identical to that of traditional publication screening. A set of search terms should be selected, and applied to published and unpublished studies. Duplicates should be excluded, and then appropriate selection criteria (e.g. study language, age of study) should be applied. This will result in a group of clinical studies that we call the *appraisal stack*.

**A2**: **extract data relating to mechanisms from these studies** Using Table 4.2, data should then be extracted from this stack of clinical studies. This will serve to identify both the content and quality of these studies. Again, we envisage that this step will accompany existing data collection protocols that are used in guideline development. Data collection should take place for each one of the reviewed studies, and a data summary table containing data summaries for each article should be produced.

**A3**: **review data for gaps** Using the completed data summary table, the analyst can then make some preliminary recommendations regarding the set of clinical research papers as a whole. These tools will particularly help to determine whether there are problems about the mechanistic aspects of this corpus of literature. We foresee several different possibilities at this stage that might require different handling.
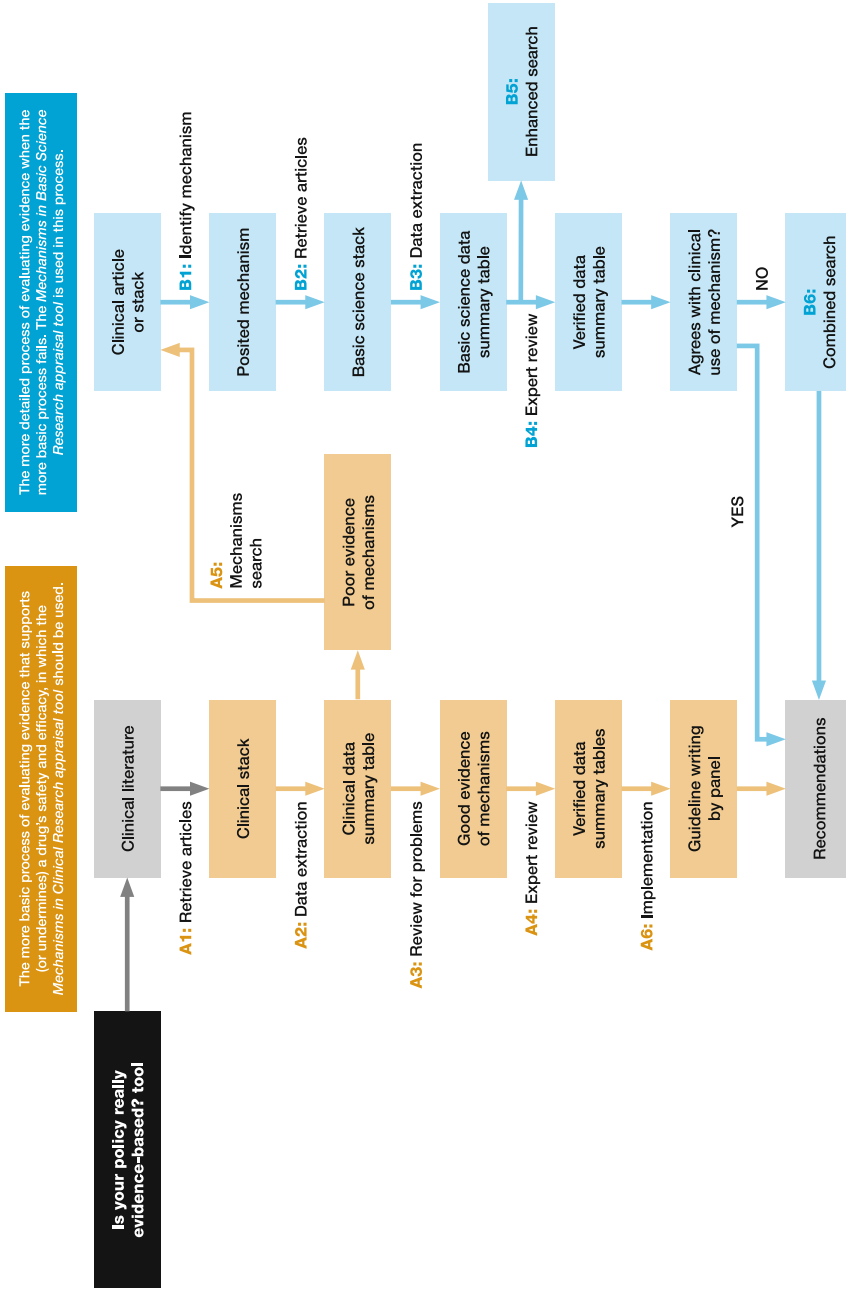
**Fig. 4.2** A suggested work-flow for the integrated use of the clinical and basic science tools

**Table 4.2** Clinical research data extraction table

| Study Details | Published? | Study Design | Mechanism Hypothesised? | Intervention used | Outcome measure | Comparison | Population |
|---|---|---|---|---|---|---|---|
| Author(s), title, summary | □ Yes – journal | □ Cluster RCT | □ No mechanism posited | Describe intervention(s) used: | Describe the outcome measure used: | Describe the comparison(s) performed: | Describe the trial population: |
| | □ Yes – other | □ Individual RCT | □ Mechanism posited: no evidence provided (please provide details) | | | | |
| | □ No | □ Correlation study | | | | | |
| | □ Unclear | □ Cross sectional study | | | | | |
| | DOI: | □ Prospective cohort study | □ Mechanism posited: evidence cited (please provide citations) | Is the choice of intervention explicitly justified? □ No □ Yes (please provide details) | Is the choice of outcome measure explicitly justified? □ No □ Yes (please provide details) | Is the comparison clinically relevant (eg. appropriate doses of comparator drugs)? □ Yes (please provide details) □ No (please provide details) | Does this differ from the intended treatment population? □ Yes (please provide details) □ No (please provide details) |
| | | □ Retrospective cohort study | | | | | |
| | Date: | □ Other, please describe | | | | | |

**Established mechanisms**: in cases where a group of clinical research papers appears to be explicitly based on a known mechanism, and where there is ample discussion of that mechanism in the basic science research literature, no further investigation will generally be required, and the user should proceed directly to stage four. A special case might be where the clinical studies appear to rely on the same mechanism, but where there is no explicit justification of that mechanism. Users in this case should make explicit note of this, and refer the issue to an expert panel (**A4**) as a possible precursor to a more developed mechanism search.

**Other cases**: in cases where the clinical research literature does not link neatly to an established mechanism, a more detailed search for a mechanism will generally be helpful to guideline authors. In this case, proceed to **A5**.

**A4**: **expert review** The data summary table should now be passed to domain experts for review. One important question at this stage is to ensure that the selection of publications examined at stage **A3** is fair and unbiased. So the experts should satisfy themselves that no cherry-picking of the research literature has taken place, and that the data extraction has fairly summarised the state of knowledge in the relevant field. If this is not the case, proceed to **A5** to conduct a more detailed mechanism search. If the domain experts are satisfied, this verified data summary table can then be passed on to a guidelines panel for use in their deliberations in **A6**.

**A5**: **mechanism search** Conduct a more detailed mechanism search using the Mechanisms in Basic Science Research Appraisal Tool to address gaps in the clinical research literature. This will frequently require consultation with domain experts for search term scoping and expert review. Once complete, the mechanisms data, together with the clinical data summary table, should be passed to an expert review panel for approval before moving to **A6**.

**A6**: **implementation/recommendation/review stage** The data summary table should then be used, in concert with other data extraction tools (and, if applicable, a summary of mechanisms data), in formulating recommendations. Here, the data summary tool is designed to facilitate panel discussions about the strengths and weaknesses of individual studies, as well as to assist with more overarching decisions about recommendations.

As discussed above, use of the *Mechanisms in Basic Science Research* tool may be necessary in some appraisals. Figure 4.2 provides an overview of the integrated use of these two tools.

## 4.4  Mechanisms in Basic Science Research Appraisal Tool

**Introduction**

This tool presents a method that a researcher would use to evaluate a mechanistic claim about a drug treatment as it appears in the basic science literature. The aim is to facilitate the construction of concise summaries for a group of basic science publications. These summaries can then be used alongside similar summaries of clinical research by a panel of experts in the context of making policy decisions. Note that this tool is not intended to produce a full reconstruction of all the mechanisms that might be relevant. Instead, the summaries will indicate the degree to which the published evidence supports some mechanism. As mechanisms frequently inform the design and interpretation of clinical trials, these summaries of evidential support for mechanistic claims that might be found in clinical research will enable a policy panel—with appropriate expert input—to appropriately evaluate both clinical and basic science research together in an integrated way.

This tool is comparatively detailed, and is therefore largely intended for use in circumstances where the details of a mechanism are particularly contentious. Broadly, this might be when either a) the consequences of a policy decision are rather serious (such as making decisions about medicines for use in pregnancy) or b) when the research base that grounds a body of clinical research is disputed or complex (such as the evaluation of treatments for chronic fatigue syndrome). Mechanisms of interest in more simple cases are likely to be dealt with adequately by our *Mechanisms in Clinical Research appraisal tool*.

**Who should use this tool**

This tool is intended for use during the development of clinical guidelines. Parts of this tool can be used by different groups as the process of guideline development proceeds. The data extraction parts of this tool (**B2** and **B3**) are likely to be largely carried out by literature review specialists alongside existing appraisal work. While these parts of the tool do assume some expertise in dealing with the medical literature, we do not assume domain-specific expertise in these parts. Parts of this tool, particularly **B1**, **B4**, and **B6**, do assume a higher level of expertise in some specified scientific domain, and this stage will largely be carried out by domain experts. Finally, **B1** and **B6** will generally require close collaboration between literature review specialists, and domain experts.

**How to use this tool**

We describe a six-stage method for using this tool. Not all steps will be necessary in each case. We generally intend this tool to follow on from issues identified during the use of the *Mechanisms in Clinical Research Appraisal Tool* (see Sect. 4.3), and this guide assumes that this is the case. Please also see the overview flowchart (Fig. 4.2) to understand the overall appraisal process.

**B1**: **identify a posited mechanism** Begin with a clinical research paper (or appraisal stack from the clinical tool). Then retrieve citations from the clinical paper(s) that describe key assumptions about mechanisms. These might include:

- Mechanism of action
- Biomarkers
- Patient population recruitment criteria
- Surrogate outcome measures

If no mechanism is described in the clinical research paper(s), or if the user is using this tool independently of the clinical research tool, expert advice is desirable at this stage to assist with the identification of a mechanism.

**B2**: **retrieve papers** Retrieve basic science papers (identified in **B1**). Then identify the purpose that these basic science papers are used for in the relevant clinical paper(s).

**B3**: **data extraction** Using Table 4.3, extract data from the relevant basic science papers identified in **B2**. Repeat for all basic science papers.

**B4**: **expert review** Pass data tables to experts for review to verify that the extraction has fairly summarised the relevant field. One important question at this stage is to ensure that the selection of publications examined at stage **B3** is fair and unbiased. Domain experts should satisfy themselves that no cherry-picking of the research literature has taken place. If extraction has not fairly summarised the field then proceed to **B5**. If however the experts are satisfied, then this verified data can then be passed to the guidelines panel for use in their deliberations. If problems and inconsistencies are revealed during this process, proceed to **B6**.

**B5**: **enhanced search (for cases where the cited literature is unrepresentative of a field)** Conduct a keyword search on the mechanism (see also Chapter 5). This should then be followed by applying stages **B1** to **B4** to the updated group of basic science papers found by this keyword search.

**B6**: **combined search (for cases where the clinical and basic sciences literature are divergent)** Conduct a combined search across both clinical and basic science material, concentrating on the connection between different kinds of evidence with respect to a claim. This will require input from experts for both the clinical and basic science material.

Once completed, the data summaries from this tool should be passed back to the relevant guideline panel, ideally in combination with the relevant clinical data summary table.

**Table 4.3** Basic science data extraction tool

| Description | Score/Reported detail | Possible red flags |
|---|---|---|
| | **Basic Information** | |
| **Short citation or DOI** | http://dx.doi.org | No DOI, predatory journal, not peer reviewed |
| **Retracted or corrected** | ☐ Not known to be retracted ☐ Retracted ☐ Corrected (please describe) | Retracted publications, substantial correction or correspondence |
| **Age of publication** | X years | Possibly outdated research, eg. papers that precede major theoretical or technological changes |
| | **Citing Study** | |
| **Does the conclusion support the citation?** | ☐ Main conclusion supports citation ☐ Conclusion does not support citation | Conclusion does not support citation |
| **Was this conclusion an intended finding of the research? Is the citation an incidental feature of the research paper? Does the conclusion arise from subgroup or post hoc analysis?** | ☐ Citation is an incidental feature of the study (please describe below) | Citation was an incidental finding of the study, raising issues of powering etc. |
| **Might there be a mismatch between the basic science research and its use in clinical research (eg. use of biomarkers, outcome measures)** | Please describe | |
| | **Types of experiment** | |
| **What is the research design?** | ☐ Experimental ☐ Observational ☐ Other/unclear ☐ Whole animal study ☐ Simulation ☐ Molecular study ☐ Cell-based ☐ Other (please describe) | Design-target mismatch or design-conclusions mismatch |
| **Are the limitations of the research explicitly discussed and evaluated?** | ☐ Evaluated ☐ Mentioned ☐ Absent | |
| **Controlled experiment?** | ☐ Yes ☐ No ☐ Unclear | |
| **Type of experimental probe?** | Please describe: | Actually or accurately described |
| **New or established experimental method?** | ☐ Established ☐ Obsolete ☐ Unclear ☐ New (please describe) | Cautions for either potentially obsolete methods, or highly novel methods without discussion of calibration and validation |
| | **Mechanism posited** | |
| **Is a mechanism described or posited in the paper? If so, what is it?** | ☐ Yes ☐ No ☐ Unclear Brief description if possible | |
| **Does the experimental finding reported bear on some aspect of that mechanism?** | ☐ Yes ☐ No ☐ Unclear | Experimental findings appear unrelated to discussed mechanism |
| **Are links drawn to relevant literature to support the mechanism described in the paper?** | ☐ Yes ☐ No ☐ Unclear | Striking absence of relevant work |
| | **Relevance to humans** | |
| **Does the paper make comments about the relevance of the research to humans?** | ☐ Yes ☐ No ☐ Unclear | Human case known to be different (eg. beta-lactam antibiotic research in rodents) |
| **Is that extrapolation explicitly justified?** | ☐ Yes ☐ No ☐ Unclear | |

## 4.5   Critical Appraisal Tool for Evidence of Mechanisms

**Introduction**

This tool presents a method for critical appraisal of mechanistic evidence which is modelled on the EBM critical appraisal worksheets publicly available at the Oxford Centre for Evidence-Based Medicine website. This aim is to provide a integrated way of evaluating the processes of gathering, evaluating, and using, evidence of mechanisms to determine the status of a causal claim. The tool is intended to be used in a stand-alone way, ideally in concert with an evaluation of other forms of evidence that might bear on a causal claim of interest. The theoretical details of these evaluations are explained in later parts of this book (see Chaps. 5, 6, and 7 respectively).

**Who should use this tool**

The tool is fairly, rather than very, detailed. It is a sensible next-step from the *Is Your Policy Really Evidence-Based tool* (Sect. 4.2) for many purposes, although we would particularly recommend it as a tool for use in contexts that are not directly related to developing healthcare guidelines. The *Mechanisms in Clinical Research appraisal tool* (Sect. 4.3) would be better fitted to these purposes (Table 4.4).

**How to use this tool**

The tool consists of eight questions. Each is accompanied with a note of guidance about both how to interpret the question (and showing how the specific question fits in with the evaluation process), as well as some notes of guidance about where to find information that will contribute to answering the question posed. Together, these questions can help reveal the strength of evidential support for some specific mechanism hypothesis.

## 4.6   GRADE-Style Tables for Mechanism Assessment

**Introduction**

One widely used approach to assessing and summarizing quality of evidence and strength of recommendations in systematic reviews and clinical practice guidelines is the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt et al. 2011), used for example by NICE (NICE 2014). The GRADE process involves collecting evidence to address a specific question about specific outcomes, and rating the quality of evidence according to the quality of study design, risk of bias, imprecision, inconsistency of findings, indirectness (relative to the target population), and magnitude of effect. The quality of evidence and strength of recommendation is then summarized in a table. GRADE tables do not include an

**Table 4.4**   A critical appraisal tool for evidence of mechanisms

| What is best? | Where do I find the information? |
|---|---|
| **What question is to be addressed with evidence of mechanisms?** ||
| The main question being addressed should be clearly formulated so that one can identify mechanistic studies relevant to the causal claim of interest. The question should clearly indicate (1) the causal claim of interest, (2) whether the question concerns efficacy or external validity, (3) the general mechanistic claim relevant to the question, and (4) one or more specific mechanism hypothesis for which evidence is sought. | The *title*, *abstract*, and *introduction* of a research article typically state the outcome and the proposed mechanism that is considered in a study. |
| **Have all the relevant mechanistic studies been identified?** ||
| The starting point for a comprehensive search for relevant studies is the major bibliographic databases (such as PubMed), but should also include a search of reference lists that appear in relevant articles, and contact with experts, particularly to enquire about unpublished studies. The search should not be limited to the English language literature. | For guidance, refer to Chapter 4 of *Evaluating evidence of mechanism in medicine*. |
| **What are the criteria used to select articles to be included in the evaluation?** ||
| The inclusion and exclusion criteria should be appropriate for the question being addressed, and should be defined in advance of the search. These criteria should specify whether the studies included have demonstrated that the effect/outcome is responsive to the proposed mechanisms, and in which way they are. For example, the criteria should specify whether the studies included demonstrated dose-dependency or not; whether *in vitro, in vivo,* or both types of study were included; or whether human, experimental animal, or both types of study were included. | The *methods* section of an article should describe the inclusion and exclusion criteria. The *abstract* and *results* sections should describe the results of the evaluation. |
| **Were the included studies sufficiently valid for the type of question asked?** ||
| One should evaluate the quality of each study by using predetermined quality criteria appropriate for the type of study. Due to the nature of mechanistic studies, these considerations vary case-by-case, but typically include assessment of biological preparation and other methods. One should consider whether the methods used can be taken to be reliable and appropriate regarding the objectives of a study. | The *methods* section of an article should describe the preparation methods, what was measured and how. Studies that do not transparently describe how the results were obtained should be excluded. |
| **Were the results similar from study to study?** ||
| Ideally, the results of different studies should be similar, but in many cases there are several different mechanism hypotheses relevant to the outcome of interest. If the latter is the case, one should evaluate which of these mechanisms are best supported by the evidence reviewed. | The *results* section of a review article should state whether the results are different from study to study. Similarly, the results section of individual studies should describe what the results are. When assessing many individual studies, one should compare the results across the studies to evaluate the concordance of the evidence. |

(continued)

**Table 4.4**   (continued)

| What is best? | Where do I find the information? |
|---|---|
| **What is the status of each specific mechanism hypothesis?** | |
| The review of evidence should determine the status of each specific mechanism hypothesis. | Considerations to take into account when evaluating the status of a mechanistic hypothesis can be found in Chapter 6 of *Evaluating evidence of mechanisms in medicine*. |
| **What is the status of the general mechanistic claim?** | |
| Once the status of each specific mechanism hypothesis has been determined, one can evaluate the status of the general mechanistic claim. | A procedure for combining evidence from mechanistic studies and evidence from clinical studies to determine the status of the general mechanistic claim is described in Chapter 6 of *Evaluating evidence of mechanisms in medicine*. |
| **What is the status of the causal claim?** | |
| Once the status of the mechanism hypothesis has been determined, one should combine the evidence of mechanisms with evidence of correlation to arrive at a status assessment for the causal claim of interest. | A procedure for combining evidence of mechanisms and correlation to evaluate the status of a causal claim is described in Chapter 7 of *Evaluating evidence of mechanisms in medicine*. |

explicit assessment of mechanistic evidence. In this tool we provide some examples of ways in which one might extend GRADE evidence profile tables to also include evidence of mechanisms. The proposed amendments are modelled according to the categories used in the GRADE tables. These amended tables illustrate that it is possible to incorporate many aspects of the approach of this book into a popular system like GRADE, without having to make any radical changes.

### GRADE-style table for mechanism assessment
### Who should use this tool
This tool is intended for use in cases where a systematic review of evidence is being conducted as part of policy development. Thus this tool is intended for a fairly expert audience, with the assumption that users will be generally familiar with current best practice in evidence appraisal. This tool is therefore an ideal step-up from the less thorough assessment that a researcher might have produced using either the *Is your policy really evidence-based?* (Sect. 4.2) and/or the *Critical appraisal tool for evidence of mechanisms* (Sect. 4.5).

### How to use this tool
Table 4.5 provides a template for an augmented GRADE-style table. We assume that a user is generally familiar with the current GRADE method for evidence appraisal. This augmented table is intended to be used a similar way. However, as it contains some questions which are likely to be unfamiliar, we have provided some notes of guidance here on these proposed new categories.

**Table 4.5** Grade table with mechanism assessment

| Interventions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Outcome:** | | | | | | | | |
| **Assessment: clinical studies** | | | | | | | | |
| No. of studies | Design | Risk of bias | Inconsistency | Indirectness | Imprecision | No. of events | Effect | Certainty |
| | | | | | | | | |
| **Outcome 2** | | | | | | | | |
| No. of studies | Design | Risk of bias | Inconsistency | Indirectness | Imprecision | No. of events | Effect | Certainty |
| | | | | | | | | |
| **Assessment: mechanism** | | | | | | | | |
| Mechanism hypothesis | Gaps | | | Masking | Inconsistency | Indirectness | Quality and status | |
| | | | | | | | | |
| **Overall assessment** | | | | | | | | |
| | | | | | | | | |

Note that providing answers to these questions may require substantial investigation, particularly in cases where the relevant mechanisms are unclear or disputed. The *Clinical Research* (Sect. 4.3) and *Basic Science* (Sect. 4.4) tools may be of value in such cases.

**Mechanism hypothesis**. If the quality of clinical studies is high, and observed effect sizes sufficiently large, there may be no need to formulate and evaluate specific mechanism hypotheses. Otherwise, each specific hypothesised mechanism should be sketched here.

**Gaps**. Crucial features of the specific mechanism hypothesis that are lacking evidence, or for which there is high risk that the available evidence is biased due to methodological limitations of the studies.

**Masking**. Evidence of mechanisms that counteract the effect of the hypothesized mechanism. This will reduce the plausibility of the intervention having a robust effect through the proposed mechanism.

**Inconsistency**. Evidence for feature(s) of a mechanism is inconsistent when there is some evidence in favour of a feature of a mechanism, and some against it, or when there is evidence for two or more mutually exclusive mechanisms. Note that inconsistency should be evaluated taking into account the amount and quality of evidence—e.g., if some of the conflicting evidence is systematically significantly less reliable due to study limitations, the inconsistency is not to be considered as severe.

**Indirectness**. Evidence relating to other populations and evidence of crucial differences between mechanisms in those populations and mechanisms in the target population.

In the **quality and status** box, one should state the overall quality of the mechanistic studies and the status of the specific mechanism hypothesis given the evidence (see Sect. 3.2 and Chap. 6). Any outstanding study limitations can be summarized here.

The **overall assessment** box should include an evaluation of the status of the general mechanistic claim, and should discuss how this informs the overall assessment of the status of the effectiveness claim. See Sect. 6.3 and Chap. 7.

**Worked example**

Table 4.6 depicts a worked example of this GRADE-style appraisal, which is an assessment of brief contact interventions for reducing self-harm. Further worked examples can be found in Appendix C.

**Table 4.6** GRADE-style table showing assessment of brief contact interventions for reducing self-harm; c.f. Milner et al. (2016)

| **Brief contact interventions** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Repeated episode of self-harm or suicide attempt** | | | | | | | | |
| **Assessment: clinical studies** | | | | | | | | |
| 12 months or less | | | | | | | | |
| No. of studies | Design | Risk of bias | Inconsistency | Indirectness | Imprecision | No. of events | Effect | Certainty |
| 11 | RCT | No serious risk of bias | No serious inconsistency | No serious indirectness | Some serious imprecision | 900 /7585 | 0.87 (0.74 to 1.04) | ⊕⊕⊕⊖Moderate |
| **Assessment: mechanism** | | | | | | | | |
| Mechanism hypothesis | | | Gaps | | Masking | Inconsistency | Indirectness | Quality and status |
| Social support: "BCIs provided participants with a sense of connectedness and the sense they were being listened to" | | | Serious gap: the link to reduction of harm | | No serious masking | No serious inconsistency | No serious indirectness | ⊕⊕⊕⊖Moderate, Status: Arguable |
| Suicide prevention literacy: "BCIs improved an individual's knowledge about suicidal behaviours or self-harm (eg, risk and protective factors), what help is available, and how to access this help" | | | Serious gap: the link to reduction of harm | | No serious masking | No serious inconsistency | No serious indirectness | ⊕⊕⊕⊖Moderate, Status: Arguable |
| Learning alternative behaviours: "BCIs involved participants learning positive and functional alternative behaviours to self-harm" | | | Serious gaps: limited evidence of the link from the intervention; the link to reduction of harm | | No serious masking | No serious inconsistency | No serious indirectness | ⊕⊕⊖⊖Low, Status: Speculative |
| **Overall assessment** | | | | | | | | |

Evidence of moderate quality suggests that brief contact interventions reduce the number of repeat episodes of self-harm or suicide attempts; however, the results are not statistically significant and the existence of a correlation can at best be considered arguable. Evidence of high quality shows brief contact interventions improve social support and suicide prevention literacy. These specific links are considered established. Evidence of moderate quality suggests brief contact interventions may also provide people with alternative coping behaviours. This specific link is arguable. However, the links from these intermediate points to the endpoint of interest, reduction of self-harm or suicide attempts, have not been investigated in detail. Overall, therefore, the specific mechanism hypotheses are arguable or speculative. Accordingly, the causal claim is at best arguable. Further research is needed to explore alternative mechanisms that could be targeted for more efficacious interventions, and to explain the lack of impact of brief contact interventions on self-harm and suicide.

## 4.7  Public Health and Social Care Tool

**Introduction**

This is a tool for appraising public health and social care policies, which differ in many ways from the kinds of interventions that are used in clinical medicine. This tool will help the authors and evaluators of these policies ensure that their interventions are as closely connected to underlying research in the relevant sectors (Fig. 4.3) as possible. Users of this tool may find the discussion of mechanisms in public health in Chap. 9 a helpful adjunct to this tool.

**Public Health and Social Care tool**
**Who should use this tool**

This tool is largely aimed at experts in public health and social care policy. It assumes a fairly high level of knowledge of the research that might be relevant for appraising a policy, and requires the user to exercise their judgement in evaluating that evidence. It is also a comparatively detailed process. A better alternative tool for contexts where a lighter review of evidence is thought to be sufficient is the *Is your policy really evidence-based?* tool found in Sect. 4.2.

**How to use this tool**

This tool can be employed as a way of checking the alignment between the available evidence of mechanisms and policy guidance. It is thus intended to help resolve problems regarding the external validity of research, and will help researchers be confident that their recommendations will be applicable to their population of interest.



**Fig. 4.3**  Our understanding of public health (after Tannahill, 1985)
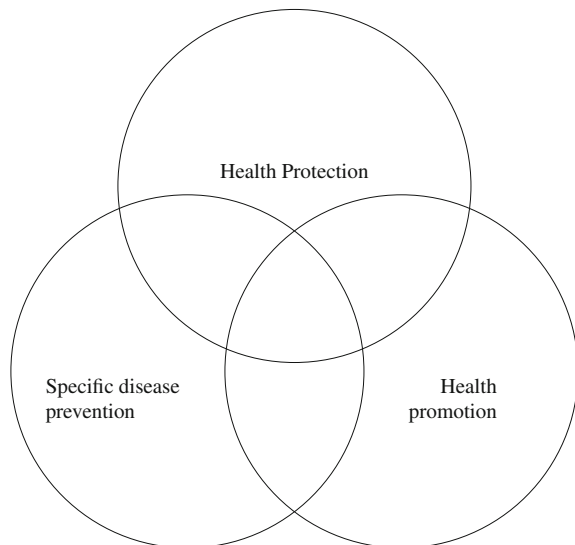
Note that the tool presupposes that population-based research (such as trials of an intervention) will be evaluated using other methods such as GRADE.

Part one of the tool (Table 4.7) asks the user to provide three sets of preliminary information: about the public health problem that the proposed intervention is intended to affect, about the nature of the intervention itself, and about the population that this intervention is meant to be applied to.

Part two of the tool (Table 4.8) then asks the user to answer questions about the evidence that bears on each of these preliminary information from part one. These questions about the evidence are divided along two axes—individual/group and biological/social. Ideally, the user should be satisfied that there are no identifiable problems in either of the four quadrants.

Note that the questions in the tools may be hard to answer in some cases. For example, research on social mechanisms may be lacking. Or, for new risks, the research base might be very slender. To offer a note of reassurance from our testing, difficulties in gathering relevant research should be regarded as a positive finding in the context of this tool.

Other parts of this book may be a helpful addition to this tool, depending on the case at hand. The *Critical Appraisal Tool for Evidence of Mechanisms* (in Sect. 4.5)

**Table 4.7**  Part one: preliminary questions for Public Health and Social Care appraisal

| What is the proposed intervention? | or | Which health outcome or behaviour is the focus of this intervention? |
|---|---|---|
|  |  |  |
| What are the desired health outcomes or behaviour changes? | or | What are the proposed interventions? |
|  |  |  |
| In what population is the intervention to be used? | | |
|  | | |

**Table 4.8** Part two: evidence questions for Public Health and Social Care appraisal

| Biological questions about groups | Social questions about groups |
|---|---|
| Which aspects of the biological mechanism(s) are known? Which are not? Why might public health interventions targeting the pathogens fail? | Which socio-economic or psychological mechanisms are known? Which are not? What is the public perception of the disease in terms of risk, seriousness and personal vulnerability? |
| Which mechanisms come into play as a population or different segments in the population react to an intervention? | |
| **Biological questions about individuals** | **Social questions about individuals** |
| Are there sub-groups within the population that should be specifically targeted? | How can behavioural mechanisms reduce exposure? |
| How can they be reached and what specific mechanisms might come into play? | |

and the *GRADE-style Tables for Mechanism Assessment* (in Sect. 4.6) would be particularly appropriate next-steps.

# References

Cooksey, D. (2006). *A review of UK health research funding*. London: The Stationery Office.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., et al. (2011). GRADE guidelines: 1. Introduction–GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*, *64*(4), 383–394.

Kingdon, J. W., & Thurber, J. A. (1984). *Agendas, alternatives, and public policies*. Boston, MA: Little, Brown.

Milner, A., Spittal, M. J., Kapur, N., Witt, K., Pirkis, J., & Carter, G. (2016). Mechanisms of brief contact interventions in clinical populations: a systematic review. *BMC Psychiatry*, *16*(1), 194.

National Audit Office (2003). *Getting the evidence: Using research in policy making*. Technical report, National Audit Office, London, HMSO.

NICE (2014). Developing NICE guidelines: The manual. Published online. https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview.

Nutley, S. M., Smith, P. C., & Davies, H. T. (2000). *What works?: Evidence-based policy and practice in public services*. Bristol: Policy Press.

Ogilvie, D., Craig, P., Griffin, S., Macintyre, S., & Wareham, N. J. (2009). A translational framework for public health research. *BMC Public Health*, *9*(1), 116.

Tannahill, A. (1985). What is health promotion? *Health education journal*, *44*(4), 167–168.