

SURE Guides for Preparing and Using Evidence-Based Policy Briefs

6. Clarifying uncertainties and needs for monitoring and evaluation

Version 2.1 – Updated November 2011

The SURE Collaboration

Suggested citation

The SURE Collaboration. SURE Guides for Preparing and Using Evidence-Based Policy Briefs: 6. Clarifying uncertainties and needs for monitoring and evaluation. Version 2.1 [updated November 2011]. The SURE Collaboration, 2011. Available from www.evipnet.org/sure

Copyright

You are free to use, distribute and reproduce this work provided the source is properly cited. You may not use this work for commercial purposes without the permission of the SURE Collaboration.

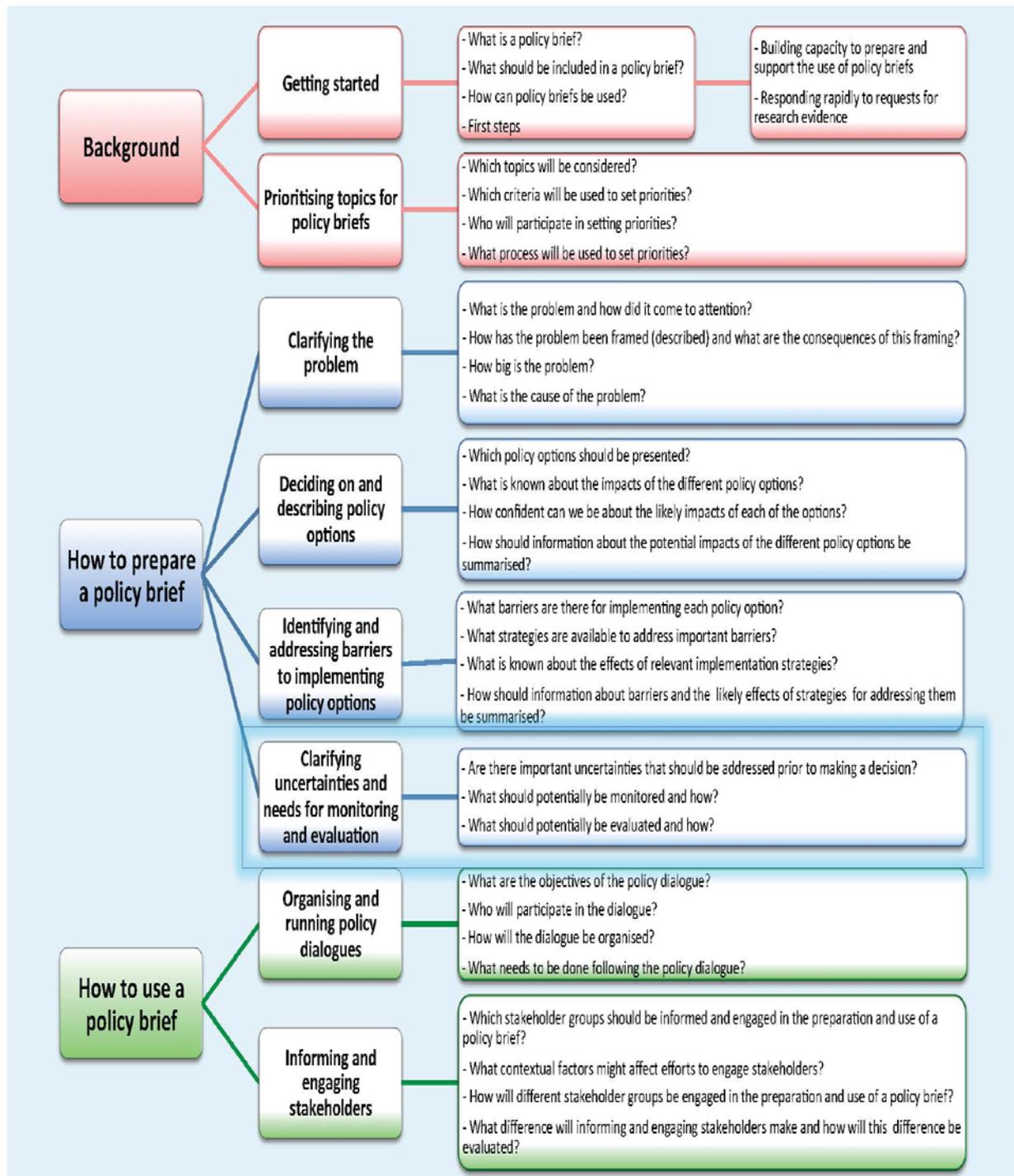


SURE is a collaborative project that builds on and supports the Evidence-Informed Policy Network (EVIPNet) in Africa and the Regional East African Community Health (REACH) Policy Initiative. The project involves teams of researchers and policymakers in seven African countries and is supported by research teams in three European countries and Canada. SURE is funded by the European Commission's 7th Framework Programme (Grant agreement no 222881).

Web address: www.evipnet.org/sure

Email address: sure.guides@nokc.no

6. Clarifying uncertainties and needs for monitoring and evaluation



Summary

Policy briefs can play an important role in ensuring that consideration is given to uncertainties about health policy options, and that such attention is given early in the policy development process rather than simply as an afterthought. The following questions can help to clarify uncertainties and needs for monitoring and evaluation:

- Are there important uncertainties that should be addressed prior to making a decision?
- What should potentially be monitored and how?
- What should potentially be evaluated and how?

The need for monitoring and evaluation should be considered at the stage when the policy options are decided (see [SURE Guide 4. Deciding on and describing policy options](#)). In this guide, the need for monitoring and evaluation is discussed in greater detail.

Evaluating the guide

As you use the guides, please complete the [evaluation form](#) included in the 'Additional resources' section so that the guide can be improved.

This page was last updated November 2011

Background

It is unrealistic to assume that we can predict the impacts of a health policy with certainty. Many – if not most – health system arrangements and implementation strategies have not been rigorously evaluated. But regardless of the availability (or paucity) of the evidence to inform decisions, policymakers must still make such choices.

Some degree of monitoring and evaluation is therefore almost always warranted (See Box 6.1), and data should be collected that can be used to:

- Reduce important uncertainties about implementing a policy
- Identify the need for corrective actions if things are not going as planned, and
- Support the continuation of the policy if things are going as planned.

Box 6.1 Monitoring and evaluation¹

Policymakers and other stakeholders will often need to know whether a new policy or programme has been implemented according to their expectations. Is the programme rollout, for example, progressing as planned? Are the objectives being achieved, and are the allocated funds being spent appropriately?

Monitoring is the term commonly used to describe the process of systematically collecting data to provide answers to such questions. The term *performance monitoring* is often used when the main focus of an evaluation is to compare how well a project, programme, or policy is being implemented, with the results that were expected.

Indicators are frequently used as part of the monitoring process. An indicator has been defined as a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance. An indicator may simply be the number of events, such as the number of vaccinations conducted within a set period of time. Or it may be a construct based on various sources of data, for example, the proportion of all children fully immunised before their first birthday.

The term *evaluation* is sometimes used interchangeably with *monitoring*, but the former usually suggests a stronger focus on the achievement of results. These terms are not used consistently and may mean different things to different people. The term *impact evaluation* is frequently used when an attempt is made to evaluate whether observed changes in outcomes (or *impacts*) can be attributed to a particular policy or programme.

Are there important uncertainties that should be addressed prior to making a decision?

While some options may be promising, they may require large investments or may be associated with important risks of adverse effects. Therefore it may be prudent to undertake an evaluation before fully implementing a policy. Substantial caution is required, for example, if interventions require large investments of scarce resources that cannot be retrieved later.² When there is important uncertainty about the benefits of interventions, a rigorous evaluation (such as a pilot study) can potentially prevent resources being wasted that could otherwise have been put to better use elsewhere. This may appear to introduce unnecessary delays, but as Julio Frenk, the former Minister of Health of Mexico notes: “Both politically, in terms of being accountable to those who fund the system, and also ethically, in terms of making sure that you make the best use possible of available resources, evaluation is absolutely critical.”³

Consideration should be given to undertaking an evaluation prior to fully implementing a policy whenever judgements about the effects of policies are based on theories, surrogate outcomes, limited observational studies, or inadequate randomised impact evaluations. The advantages and disadvantages of undertaking an evaluation prior to the full implementation of a policy are summarised in [Table 6.1](#). In general, whenever there is a substantial risk of the benefits of a policy not outweighing the harms and costs, or a potential need to make important modifications to a policy, the advantages of an evaluation are likely to outweigh the disadvantages. However, this depends on whether an evaluation can be completed within an acceptable timeframe, and whether the evaluation is feasible and affordable.

In a policy brief there are several ways to outline options that have important uncertainties, which should be addressed prior to full implementation. Depending on how compelling the arguments are for undertaking an evaluation, these options can include:

- Flagging the uncertainties and ensuring that consideration is given to undertaking an evaluation before the full implementation
- Describing the full implementation and undertaking an evaluation as two separate options (see the example link below)
- Describing the option with an evaluation being undertaken first as part of the option (in other words, the option should not be fully implemented without first undertaking an evaluation)

In an [example](#) of a policy brief on integrating mental health into primary care in Zambia, two options were offered: one was full implementation and the other was incremental implementation with an initial evaluation. Further details about this study can be found in the [‘Additional resources’](#) section of this guide.

Table 6.1 Advantages and disadvantages of undertaking an evaluation prior to deciding whether to implement a policy

	Findings of the evaluation	
	The balance between the benefits, harms and costs	
	Favour the policy	Do not favour the policy
Advantages	<ul style="list-style-type: none"> • Potential for improvements prior to implementation 	<ul style="list-style-type: none"> • Possible to stop implementation and to reconsider options
Disadvantages	<ul style="list-style-type: none"> • Delay in implementation 	

This page was last updated November 2011

What should potentially be monitored and how?

Consideration should be given to the need for monitoring when options are described in a policy brief. The extent to which monitoring is necessary and what exactly should be monitored will depend on how much uncertainty there is regarding the inputs, activities, outputs, and impacts of an option (as summarised in Table 6.2.^{1,4}). The extent to which specific types of uncertainty should be described in a policy brief will depend on a number of factors, including the:

- Degree of uncertainty
- Potential for monitoring to reduce important uncertainties
- Feasibility of monitoring
- Ability and preparedness to act on the results of monitoring

Table 6.2 Types of indicators, reasons for monitoring them, and potential actions*

Types of indicators	Definitions	Reasons for monitoring	Potential actions
Inputs	Financial, human, and material resources	Uncertainty about the magnitude of the resources required	Adjustments to the budget
Activities	Ways in which inputs are used	Uncertainty about how resources will be used, or to ensure that resources are used as intended	Changes in how resources are allocated or used
Outputs	Services provided, changes or short-term effects of activities	Uncertainty about the immediate effects of activities	Changes in the budget, how resources are used and the activities that are undertaken, or a decision to continue or discontinue the implementation of the option
Impacts	Desirable and undesirable effects that are important to those affected	Uncertainty about the extent to which outputs accurately and comprehensively reflect the likely impacts of the option	Changes in the services provided or their provision, or a decision to continue or discontinue the implementation of the option, or the undertaking of an impact evaluation

*Adapted from references 1 and □

When there are compelling reasons for monitoring, consideration should be given to which specific indicators to use, in other words, which specific types of data could potentially be collected to provide a reliable measure of inputs, activities, outputs, or impacts. The data may simply be a record of the number of events, such as how many vaccinations were conducted within a set period of time. Or the data may be constructs based on various data sources, such as the proportion of all children who are fully immunised before their first birthday. The choice of indicators to be used for monitoring should be based on their specific characteristics, including the extent to which:

- They are acceptable to those who are being assessed and those undertaking the assessment
- Valid, reliable, and consistent data are available for collection
- They are affordable

Consideration of which indicators to use and how to incorporate these considerations in a policy brief will help to ensure that important uncertainties are recognised, as well as appropriate ways of addressing them. This will also help to inform discussions about the need for monitoring an option during the policy development process. Although the amount of detail related to monitoring that can be included in a policy brief is limited, the risk of not incorporating these considerations is that monitoring may not be seen as an integral component of the option itself, and may be included only as an afterthought. Key related considerations could be incorporated into the text of the full policy brief report; more detailed considerations, where relevant, could be included as an appendix. A [worksheet](#) to help with the systematic and transparent consideration of the need for monitoring is provided in the '[Additional resources](#)' section of this guide. Further guidance is available in the [SUPPORT Tool on planning the monitoring and evaluation of policies](#). A SURE Rapid Response [article on policy implementation monitoring](#) is also provided.

This page was last updated November 2011

What should potentially be evaluated and how?

Although monitoring can sometimes be used to reduce uncertainties about the impacts of an option, on its own it may not be adequate, and an evaluation may be necessary. Monitoring does not necessarily indicate whether an option has had an impact on the indicators measured: these will almost always be influenced by factors other than the actual option that is being implemented. This makes it extremely difficult to determine what may have caused the changes that are observed. For example, although monitoring may reveal a performance improvement over time, the actual implementation of the option may not be the only causal factor or may not have caused the change at all.

Evaluation of an option should, as far as possible, measure all outcomes, both desirable and undesirable, for which there is substantial uncertainty and which are important to those affected (i.e. where the results of the evaluation could conceivably affect a decision about whether the implementation of an option is worthwhile).

An impact evaluation must estimate what would normally occur in the absence of implementing an option, and then compare this to an estimate of what happens when an option is implemented. Ideally, an evaluation should be built into a programme during the design phase to ensure that the evaluation is planned and implemented as early as possible. Policy briefs can help to ensure that appropriate consideration is given to the need for impact evaluations and to when and how they should be conducted.

It is important that evaluation methods and findings are as reliable as possible. Attributing an observed change to the implementation of an option requires a comparison between the individuals or groups exposed to the option and those who are not. The groups compared should be as similar as possible in order to ensure that only those influences related to the option are evaluated and not others. The most effective way to do this is to conduct a randomised trial in which individuals or groups of people (e.g. within specific geographic areas) are randomly allocated to either receive the option or not to receive.^{5,6,7,8} Randomised trials can be conducted as pilot projects before a programme is introduced at a national level, or they can be undertaken in parallel with full scale implementation, for example, by randomly allocating the districts in which an option will be implemented first and then comparing the results to other districts where the implementation has been delayed.

Randomised trials, however, may not always be feasible. Alternative approaches include interrupted time series analyses and controlled before-after studies.⁹ An interrupted time series analysis can be used when data are collected from multiple time points both before and after the implementation of the option. If the necessary data are available, interrupted time series analyses are relatively easy to conduct and are advantageous both because a control group is not needed and because their design controls for trends over time and variability in indicators over time. The most important disadvantage of such analyses, however, is that influences other than the option that is being evaluated may also affect the observed changes.

In controlled before-after studies, changes before and after the implementation of an option are compared with the changes observed during the same time period in areas where the option has not been implemented (e.g. in neighbouring districts or countries).

The main advantage of controlled before-after studies is that they may sometimes be the only practical option in situations where, for example, randomisation is not feasible for practical or political reasons and where it is not feasible to collect data at multiple time points. However, controlled before-after studies rarely provide reliable estimates of impacts. This is because known or unknown differences between the compared groups may exert more influence on the outcomes measured than the actual option that is being evaluated. Consequently, it is generally difficult – if not impossible – to attribute with confidence any observed changes (or lack of change) to the implementation of an option.

Other study designs may sometimes be used to assess the impacts of health policy options. However these are often not feasible for assessing the impacts of health policies (e.g. cohort studies and case-control studies) and they rarely provide compelling evidence (particularly before-after studies, historically controlled studies and cross-sectional studies).^{1,8} Although qualitative studies (as well as other quantitative designs such as surveys) can provide valuable evidence to explain how an option worked or why it did or did not work, beyond gathering the perceptions of those who were interviewed or surveyed, they are unable to generate the kind of data that can be used to estimate the effect of an option.

A [list of different evaluation designs](#) with definitions and a summary of their [strengths and weaknesses](#) are included in the 'Additional resources' section of this guide. Further guidance is available in the [SUPPORT Tool on planning the monitoring and evaluation of policies](#).

Rigorous impact evaluations can be expensive and budget, time, or data constraints may severely limit the ability to undertake them. These constraints can reduce the reliability of impact evaluations because:

- The overall validity of the results may be threatened by a number of factors including insufficient planning or follow-up, a lack of baseline data, a reliance on inadequate data sources, or the selection of an inappropriate comparison
- Samples may be inadequate such as convenience samples that are not representative, samples that are too small, or if inadequate attention is given to contextual factors

It may be possible to address budget, time, and data constraints by, starting the planning process early and reducing the cost of data collection. However, for an impact evaluation to be worthwhile, it is important to ensure that neither the threats to the validity of the results nor the limitations of the sample are such that the results of the evaluation fail to provide reliable information. Before implementing an evaluation, an assessment should therefore be made as to whether an adequate evaluation is possible. If it is not, an assessment should be made as to whether the programme should be implemented without evaluation, given the nature of the uncertainty about its potential impacts.

Several models have been described for assessing the extent to which an adequate evaluation is possible and this is sometimes referred to as an “evaluability assessment”.¹⁰ An evaluability assessment can help to determine whether the intended objectives of an evaluation can be achieved with the resources and data available and within the specified time horizon of the evaluation.¹¹ The purpose is not to see if the logic of a programme is sufficiently clear for an evaluation design to be constructed, but rather to check if the particular level or levels of government (or non-governmental organisations) are able to begin collecting, analysing, and reporting evaluation data so as to inform the decision-making process.

A policy option is evaluable if:

- The goals and priority information needs are well defined
- The goals are plausible
- Relevant data can be obtained at a reasonable cost, and
- The intended users of the evaluation results are able to agree on how they will use the information¹²

A variety of methods can be used to assess the evaluability of a programme, including interviews, document reviews, and site visits.⁹ A [worksheet for considering needs and alternatives for evaluating the impacts of an option](#) is provided in the ‘Additional resources’ section of this guide.

Decisions about whether to proceed with an option when there are important uncertainties about its impacts and it is not evaluable will depend on judgements about the size of the problem, what the alternatives are, the expected impacts of the programme, and the extent of uncertainty about those impacts. They will also depend on the value, costs, feasibility, and acceptability of the option.^{13, 14} The [SUPPORT Tool on dealing with insufficient research evidence](#) is provided in the ‘Additional resources’ section of this guide.

[Workshop materials and a presentation on clarifying uncertainties and needs for monitoring and evaluation](#) are available in the ‘Additional resources’ section of this guide.

This page was last updated November 2011

Additional resources

[Evaluation form](#)

A form for evaluating the SURE Guides

[Glossary](#)

A glossary of terms used in the guides

[Example of evaluation of implementation as one of policy options](#)

Example of a policy brief which presented two options: one being full implementation and the other being incremental implementation (in which an evaluation is completed first)

[Worksheet for considering needs and alternatives for monitoring implementation of an option](#)

Worksheet for considering the need and alternatives for monitoring the implementation of an option

[SUPPORT Tool for monitoring and evaluation](#)

Questions to consider when planning monitoring and evaluation of policy options

[A SURE Rapid Response article on policy implementation monitoring](#)

An article summarising what policy implementation monitoring entails

[Evaluation designs](#)

A list of evaluation designs and definitions from the SUPPORT Tool for planning monitoring and evaluation of policies

[Strengths and weaknesses of evaluation designs](#)

A summary of the strengths and weaknesses of evaluation designs from the SUPPORT Tool for planning monitoring and evaluation

[Worksheet to assist with assessing the need for evaluating the impacts of an option](#)

Worksheet for considering the need and alternatives for evaluating the impacts of an option

[SUPPORT Tool for insufficient evidence](#)

Questions to consider when dealing with insufficient research evidence

[Workshop materials and presentations](#)

Guides for a workshop and a PowerPoint presentation on clarifying uncertainties and needs for monitoring and evaluation

This page was last updated November 2011

References

1. Fretheim A, Oxman AD, Lavis JN, Lewin S. SUPPORT Tools for evidence-informed health Policymaking (STP): 18. Planning monitoring and evaluation of policies. *Health Res Policy Syst* 2009, 7(Suppl 1):S18.
2. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *J R Soc Med* 2007; 100:453-60.
<http://jrsm.rsmjournals.com/cgi/content/full/100/10/453>
3. Moynihan R, Oxman AD, Lavis JN, Paulsen E. Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. Rapport Nr 1-2008. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2008.
4. Development Assistance Committee Working Party on Aid Evaluation: Glossary of Key Terms in Evaluation and Results Based Management. Paris: OECD Publications, 2002. www.oecd.org/dataoecd/29/21/2754804.pdf
5. Kleijnen J, Gøtzsche P, Kunz RH, Oxman AD, Chalmers I. So what's so special about randomisation? In: Maynard A, Chalmers I., editor(s). *Non-random Reflections on Health Services Research: on the 25th Anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Books, 1997:93-106.
6. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317(7167):1185-90.
7. Savedoff WD, Levine R, Birdsall N. When will we ever learn? Improving lives through impact evaluation. Washington, DC: Center for Global Development, 2006. www.cgdev.org/files/7973_file_WillWeEverLearn.pdf
8. Oliver S, Bagnall AM, Thomas J, Shepherd J, Sowden A, White I, et al. Randomised controlled trials for policy interventions: a review of reviews and meta-regression. *Health Technol Assess* 2010;14(16).
9. Eccles M, Grimshaw J, Campbell M, et al. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 2003; 12:47-52.
10. Trevisan MS. Evaluability assessment from 1986 to 2006. *American Journal of Evaluation* 2007; 28:290-303.
11. Kusek JZ, Rist R. Ten step to a results based monitoring and evaluation systems. In: Segone M, ed. *Bridging the gap: The role of monitoring and evaluation in evidence-based policy making*. UNICEF Evaluation Working Papers. Issue 12, 2008, 98-119. www.unicef.org/ceecis/evidence_based_policy_making.pdf

12. Wholey JS. Evaluability assessment. In: Wholey JS, Hatry HP, Newcomer KE, eds. Handbook of Practical Program Evaluation. Second edition. Hoboken, NJ: John Wiley and Sons, 2004, 33-62.
13. Oxman AD, Lavis JN, Fretheim A, Lewin S: SUPPORT Tools for evidence-informed health Policymaking (STP). 16. Using research evidence in balancing the pros and cons of policies. [Health Res Policy Syst. 2009, 7\(Suppl 1\):S16.](#)
14. Oxman AD, Lavis JN, Fretheim A, Lewin S: SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence. [Health Res Policy Syst. 2009, 7\(Suppl 1\):S17.](#)

This page was last updated November 2011