DRAFT: 2009 06 12

SUPPORT Tools for evidence-informed health Policymaking (STP)

12. Dealing with insufficient evidence

Andrew D Oxman[1]
John N Lavis[2]
Atle Fretheim[3]
Simon Lewin[4]

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway. Email: oxman@online.no
2. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5. Email: lavisj@mcmaster.ca
3. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway. Email: atle.fretheim@nokc.no
4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and Health Systems Research Unit, Medical Research Council of South Africa. Email: simon.lewin@nokc.no

Corresponding author:
Dr Andy Oxman
Norwegian Knowledge Centre for the Health Services
P.O. Box 7004, St. Olavs plass
N-0130 Oslo, Norway

Email: oxman@online.no

Abstract

Background: This article is number 12 in a series of 21 articles on tools for evidence-informed health policymaking and addresses the issue of decision making when there is insufficient evidence. Policymakers often have insufficient evidence to know with certainty what the impacts of a health policy or programme will be. They must *still* make decisions.

Objectives: In this article, we suggest four questions that can be considered when there is insufficient evidence to be confident about the impacts of implementing a policy or programme.

Key messages:

- If there is insufficient evidence to allow one to be confident about the impacts of implementing a policy or programme, the following questions should be considered:
  1. Is there a systematic review of the impacts of the policy or programme?
  2. Has inconclusive evidence been misinterpreted as evidence of no effect?
  3. Is it possible to be confident about a decision despite a lack of evidence?
  4. Is the policy or programme potentially harmful, ineffective or not worth the cost?
- Common mistakes that should be avoided include:
  - Making assumptions about the evidence without a systematic review
  - Confusing a lack of evidence with evidence of no effect
  - Assuming that insufficient evidence necessarily implies uncertainty about a decision
  - Assuming that it is politically expedient to feign certainty, and
- Saying “no” to a policy or programme, or saying “yes” when there are important uncertainties can have undesirable consequences. Often it would be better to say “only with a well-designed impact evaluation”

Background

This article is number 12 in a series of 21 articles on tools for evidence-informed health policymaking [1]. It is also the sixth of six articles in the series that characterise the costs and consequences of potential policy and programme options. In this article we suggest four questions that can be considered when there is insufficient evidence to inform judgements about the impacts of policy options.

It is rare to know with certainty what the impacts of a health policy or programme will be. Many governance, financial and delivery arrangements have not been rigorously evaluated, and neither have many of the programmes, services and drugs that these arrangements support. Therefore there is often a paucity of evidence upon which to base decisions. But policymakers must still make decisions, regardless of the availability of evidence to inform those decisions.

In this section, we focus on decision making when there is insufficient evidence to know whether a policy or programme will have intended impacts or whether it will have unintended (and undesirable) impacts. We suggest four questions that can help to avoid the following mistakes that are commonly made in relation to insufficient evidence:
making assumptions about the evidence without a systematic review, confusing a lack of evidence with evidence of no effect, assuming that insufficient evidence necessarily implies uncertainty about a decision, and the assumption that it is politically expedient to feign certainty.

Questions to consider

If there is insufficient evidence to allow one to be confident about the impacts of implementing a policy or programme, the following questions can be considered:
1. Is there a systematic review of the impacts of the policy or programme?
2. Has inconclusive evidence been misinterpreted as evidence of no effect?
3. Is it possible to be confident about a decision despite there being a lack of evidence?
4. Is the policy or programme potentially harmful, ineffective or not worth the cost?

1. Is there a systematic review of the impacts of the policy or programme?

The first step in addressing a potential lack of evidence is to find out what evidence is available. It is risky to make assumptions about the availability of evidence without having a systematic review. In this series we address considerations related to finding and critically appraising systematic reviews in Articles 5 and 6 [2, 3].

For many health systems questions it may not be possible to find relevant and up-to-date systematic reviews. For example, despite widespread recognition that health workers are critical to achieving the MDGs and other health goals, an overview of systematic reviews of policy options to address human resources for health, found only a small amount of high-quality, synthesised research evidence regarding the effects of a few policy options for the improvement of human resources for health [4]. Other overviews of reviews have found similar gaps. The lack of a systematic review does not necessarily reflect a lack of evidence,

but it is difficult for policymakers to know what evidence there is under such circumstances (see Box 1 for example).

When time or resources are limited it may be necessary to undertake rapid assessments. These assessments should be transparent about the methods used, as well as any important limitations of these methods or related uncertainties. They should also address the need for, and urgency of, undertaking a full systematic review [5]. Consideration should also be given to commissioning a new review whenever a relevant, up-to-date review of good quality is found to be unavailable. Appropriate processes for doing this, including processes for setting priorities for systematic reviews, should also be assessed. Building on and strengthening international collaborations such as the Cochrane Collaboration (www.cochrane.org).can help to avoid unnecessary duplication of efforts in producing systematic reviews and can help to ensure that up-to-date reviews are more readily available.

2. Has inconclusive evidence been misinterpreted as evidence of no effect?

A common mistake in instances when there is inconclusive evidence, is the confusion between 'no evidence of an effect' and 'evidence of no effect' [6]. When there is inconclusive evidence, it is wrong to claim that it shows that a policy or programme has had 'no effect'. 'Statistical significance' should be interpreted cautiously and should *not* be confused with importance.

When results are 'statistically non-significant' it means that the null hypothesis (i.e. that there was no impact) cannot be rejected. Typically a cut off of 5% is used: in other words, if the statistical analysis shows that impacts as large as, or larger than, the observed effect would be expected to occur more than one out of twenty times by chance ($p \geq 0.05$), the results are considered to be 'statistically non-significant'. There are, however, two problems with this assumption. Firstly, the cut-off point of 5% is arbitrary. Secondly, 'statistically non-significant' results, which are often mislabelled as 'negative' might or might not be inconclusive (see Box 2).

It may be equally misleading to describe a 'positive' but statistically non-significant trend (suggesting a possible beneficial impact) as 'promising'. This is done commonly, whereas a 'negative' effect of the same magnitude (suggesting a possible harmful impact) will not be typically described as a 'warning sign'.

Policymakers should be aware that researchers commonly make both of these mistakes, and to avoid being misled they should be cautious about possible misinterpretations of statistical significance.

3. Is it possible to be confident about a decision despite a lack of evidence?

While some policymakers would agree with Charlie Brown who said, "I am always certain if it is a matter of opinion," most would agree that high quality evidence provides a better basis for being confident about a decision. Nonetheless, there may be good reasons for being confident about a decision even when there is a lack of evidence. For example, there is very low quality evidence that giving aspirin to children with influenza or chicken pox may cause Reye's syndrome (a rare but deadly condition) [7]. Despite the limitations of this evidence,

the US Surgeon General and others have confidently advised against the use of aspirin in these circumstances. This is because of the availability of paracetamol (acetaminophen) as an equally effective and inexpensive alternative which allows children not to be put at risk, even if there is uncertainty about the actual level of the risk itself. Conversely, it may be reasonable to be confident that policies or programmes with high costs and potentially serious adverse effects should *not* be rolled out without a rigorous impact evaluation.

4. Is the policy or programme potentially harmful, ineffective or not worth the cost?

It is risky not to acknowledge uncertainty for the sake of political expediency. As we noted in the introduction to this series (see Article 1) [1], acknowledging that there is imperfect information to inform policies can reduce political risk because it allows policymakers to set in motion ways to alter course if policies do not work as expected. There is far greater political risk when policymakers confidently advocate a policy for which there is a lack of evidence. Adhering to such programme regardless of the results, may subject the programmes to political criticism for unrelated or uncontrollable contingencies.

Good intentions and plausible theories are insufficient as a basis for selecting policies and practices (see Box 3). This is true for health systems as well as clinical interventions. Examples of clinical interventions that were believed to be beneficial, were widely used, and then subsequently found to be relatively ineffective or harmful include:

- High instead of low osmolar rehydration solutions for children with diarrhoea [8]
- Diazepam or phenytoin instead of magnesium sulphate for women with eclampsia [9, 10]
- Six or more antenatal care visits instead of four [11]
- Corticosteroids for patients with severe head trauma [12]
- Albumin instead of salt water for resuscitation in critically ill patients [13]
- Hormone replacement therapy to reduce the risk of coronary heart disease in women [14]
- Electronic mosquito repellents for preventing mosquito bites and malaria infection [15]

All of the above interventions were based on underlying theories, indirect evidence, and were typically from surrogate outcomes, and observational studies. Subsequent randomised trials disproved the prior assumptions, supporting the assertion (quoted in Box 3) that the public can be served more responsibly and ethically by making rigorous evaluations an expectation rather than an option for informing decisions related to clinical interventions.

These same issues are true for health systems and public health interventions. Examples of health systems and public health interventions that have been widely used and advocated, which may be ineffective, and may do more harm than good, include:

- Educational and community interventions to reduce the risk of teenage pregnancy [16]
- Directly observed therapy for tuberculosis [17]
- User fees for essential medicines [18]
- For-profit instead of not-for-profit private hospitals [19]
- Reducing maldistribution by requiring doctors to spend a minimum number of years in an underserved area before allowing them to specialise [20]
- Some forms of results-based financing or pay-for-performance [21]
- Contracting with the private sector to provide health services [22]

Substantial caution is required before investing scarce resources in policies or programmes that require large investments that cannot be retrieved [23]. If there is important uncertainty

about the impacts of such policies or programmes, a rigorous evaluation such as a pilot study, for example, can potentially prevent wasting resources that could be better spent elsewhere. And while such undertakings may appear to be unnecessary delays, Julio Frenk, the former Minister of Health of Mexico, has noted: “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” [24]. Both saying “no” to a programme or saying “yes” can have undesirable consequences (see Box 4). A better approach is to inform policy making by testing the proposed intervention within a well-designed impact evaluation.

Whenever judgements about the effects of policies or programmes are based on theories, surrogate outcomes, limited observational studies or inadequate impact evaluations, policymakers should be cautious about implementing them (see Box 5 for example) [25].

Even when there is little uncertainty about the benefits of a policy or programme, there may still be important uncertainty about other potentially important consequences, including unintended effects (harms) and costs (see Box 6 for example). Health system and public health interventions can cause harm, including even policies or programmes with compelling rationales.

For a policy or programme that is promising, but for which there is insufficient evidence to be confident about whether it is potentially harmful, ineffective, or not worth the cost, consideration should be given to demanding a well-designed impact evaluation either prior to rolling out the policy or programme, or integrated as part of the rollout. We address further considerations regarding monitoring and evaluation in Article 15 of this series [25].

Resources

Useful documents and further reading

- 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

- Oxman AD, Bjørndal A, Becerra F, Gonzalez Block MA, Haines A, Hooker Odom C, et al. Helping to ensure well-informed public policy decisions: a framework for mandatory impact evaluation. Lancet 2009; in press.

Box 1. An independent inquiry into inequalities in health – an example of the need for up-to-date systematic reviews to know what evidence there is

In 1997, the incoming British Labour government was keen to reduce inequalities in health. To do this, it set about obtaining advice from the public health community about how to reduce inequalities, but set clear limits about what advice it would find acceptable. The government wanted the advice quickly but stipulated that the advice had to be backed by evidence, in keeping with the government's expressed desire that public policy should be based on evidence [26]. The public health and other communities responded enthusiastically and considerable material of interest was produced by, and for, the inquiry, and many recommendations made [27].

Subsequent reviews of those recommendations, however, found little evidence for the likely or actual effectiveness of many of the recommendations made [27]. There was also a striking lack of adequate searches for relevant evidence or attempts to avoid bias in the way information was identified, appraised and used.

This is not to suggest that governments cannot develop or implement policies that lack the support of unequivocal evidence. However, a lack of evidence does make it difficult for them to decide on priorities. The readiness of researchers to recommend policies when they know little about the likely effectiveness, makes this more difficult still.

The task of this particular inquiry in the United Kingdom would have been easier if up-to-date systematic reviews had been available. Further, a system to ensure that the inquiry's recommendations would be reviewed regularly as new information emerged from updated systematic reviews would have helped to ensure that adjustments in policies could have been made as new evidence became available. It could also have helped to avoid similar future difficulties when similar inquiries were undertaken or similar policies considered in other jurisdictions. The Cochrane Collaboration (www.cochrane.org) and the Campbell Collaboration (http://www.campbellcollaboration.org/) are examples of international networks that have structures for preparing and keeping up-to-date systematic reviews which can facilitate this.

The investment of public resources in primary research has been substantial and remains so, but the returns remain far less than they might otherwise be, and the results scattered rather than synthesised. People faced with tasks and timescales similar to those of the British inquiry would be assisted greatly if up-to-date systematic reviews were more readily available. There are no unequivocal answers to the question "What works?" in terms of developing health policies and programmes, but a systematic review is the best starting point for finding out what is known.

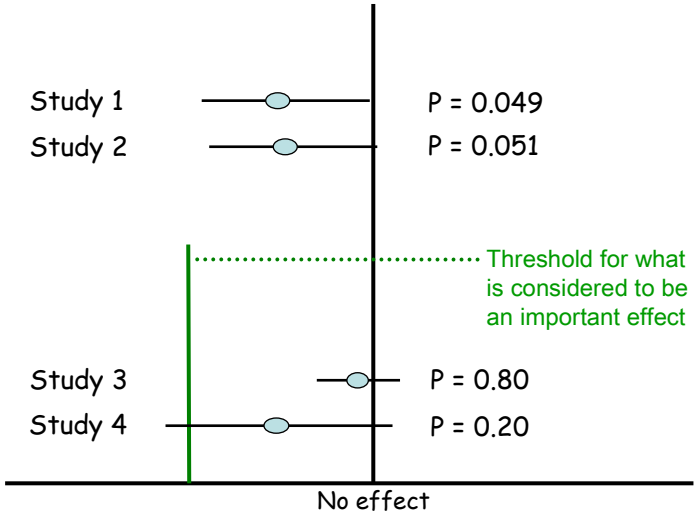
Box 2. "Statistical non-significance"

Figure 1 below illustrates two problems that arise when results are classified as 'statistically non-significant' or 'negative':

1. *The classification is based on an arbitrary cut-off*. The results from Study 1, for example, are marginally different from the results from Study 2. But by using the conventional cut-off of $P < 0.05$, the results for Study 1 are ranked as 'statistically significant' and the results for Study 2 as 'statistically non-significant'.

2. *'Statistically non-significant' results may or may not be inconclusive*. If the green short vertical line in the figure below indicates the smallest effect considered important, the results for Study 3 would be conclusive, since an important impact is highly unlikely, while, the results for Study 4 would be categorised as 'inconclusive', since it is not unlikely that there would be an important impact (the 95% confidence interval crosses the threshold for what was considered to be an important effect). Both results, however, might be regarded as 'statistically non-significant' or 'negative'.

Figure 1



The dots (⬭) in the Figure above indicate the estimated effect for each study and the horizontal lines indicate the 95% confidence intervals. A 95% confidence interval means that we can be 95 percent confident that the true size of effect is between the lower and upper confidence limit (the two ends of the horizontal lines). Conversely, there is a 5%t chance that the true effect is outside this range.

Box 3. Good intentions and plausible theories are insufficient

"Professional good intentions and plausible theories are insufficient for selecting policies and practices for protecting, promoting and restoring health. Humility and uncertainty are preconditions for unbiased assessments of the effects of the prescriptions and proscriptions of policy makers and practitioners for other people. We will serve the public more responsibly and ethically when research designed to reduce the likelihood that we will be misled by bias and the play of chance has become an expected element of professional and policy making practice, not an optional add-on." (Ian Chalmers, Editor, the James Lind Library, 2003)[*] [28].

[*] From a presentation at the Norwegian Directorate for Health and Social Welfare, 1 September 2003. For a more detailed discussion of these comments see Reference **Error! Bookmark not defined.**.

Box 4. The consequences of saying “no” or “yes” instead of “only in research”

High-income countries, as well as low- and middle-income countries, have limited resources. For this reason, in the United Kingdom for example, the National Institute for Health and Clinical Excellence (NICE) officially recognises the principle of recommending that when important uncertainties exist about an intervention’s effects, such interventions should only be used in the context of research [23]. Sixteen (approximately 4%) of NICE’s technology appraisal recommendations published between 1999 and early 2007 advised the use of a technology only in the context of research. The consequences of getting decisions wrong by either saying “no” or “yes” to a technology without doing this, are summarised below (see Reference **Error! Bookmark not defined.** for further details).

The consequence of saying “no” instead of “only in research”

- Patients are denied access to promising and potentially effective technologies
- There are delays in building the evidence base in key areas, with a resulting negative overall impact on health outcomes

The consequences of saying “yes” instead of “only in research”

- Access to unproven and potentially harmful or ineffective interventions is promoted
- Any ongoing or future research in the field is severely hindered. Important questions on effectiveness and cost-effectiveness may never be answered
- Limited resources are wasted
- Having to reverse a yes decision in light of any future evidence compromises credibility and is difficult to implement

Box 5. An example of a potentially ineffective or harmful intervention that has been widely promoted based on insufficient evidence

Effective drugs for tuberculosis have been available since the 1940s, but two million people continue to die from the disease each year, mostly in low-income countries. People with tuberculosis require treatment for at least six to eight months. Many find it difficult to complete their course of treatment and this serves as a major constraint to eradicating the disease. Poor adherence to treatment can lead to prolonged infectiousness, drug resistance, relapses, or even death. Incomplete treatment thus poses a serious risk both to the individual and to communities as a whole.

Directly observed therapy (DOT) seeks to improve the adherence of people to tuberculosis treatment by using health workers, family members, or community members to directly observe patients taking their anti-tuberculosis drugs. DOT is potentially advantageous because adherence may improve when people are closely monitored and there is a social process involving peer pressure. Potential disadvantages include the fact that this treatment moves away from adherence models of communication, with their emphasis on cooperation between patient and provider, back to a traditional medical approach where the patient is a passive recipient of advice and treatment. Resource implications for such a policy are substantial, particularly in low- and middle-income countries where the case load may be high. DOT may also make adherence worse if it is rigidly applied in an authoritarian setting, or where people are expected to travel considerable distances to have their treatment supervised.

The World Health Organization (WHO) and others have actively promoted DOT since the 1980s, generally as part of a comprehensive tuberculosis management programme known as DOTS (directly observed therapy, short course), a five-element strategy for the control of tuberculosis. Although the strategy as a whole appears sound, there is substantial uncertainty about DOT as a key element of DOTS. When DOTS was originally launched, the evidence for the effectiveness of DOT came entirely from observational studies and no randomised impact evaluations of DOT had been undertaken. Subsequently, eleven randomised trials have compared DOT with self-administration and found that DOT did not improve adherence, despite the substantial resources required and its other disadvantages [17].

Box 6. An example of important uncertainties about potentially important harms

Although there is little doubt that financial incentives, if they are large enough, can change behaviours, they can also cause unintended behaviours, and the costs, including both the incentives themselves and the cost of administering them, can be substantial [21]. Unintended effects of paying for performance (the provision of payment for the attainment of well-defined results) that have been observed include:

- **Unintended behaviours**
  Conditional cash transfers (CCT) caused some mothers to keep their children malnourished in order to retain eligibility. CCT may also have increased fertility by 2% to 4% in another study, because only pregnant women were eligible for a subsidy
- **Distortions**
  Financial incentives may cause recipients to ignore other important tasks
- **Gaming**
  Financial incentives can result in gaming (changes in reporting rather than desired changes in practice)
- **Corruption**
  Financial incentives may be stolen or misused, if not adequately managed
- **Cherry-picking**
  Performance incentives for providers can influence whether health care is accessible to patients by altering how willing health care workers or organisations are to care for sicker patients, more disadvantaged populations, or more difficult patients
- **Widening the resource gap between rich and poor**
  Performance incentives for providers may widen the resource gap that exists between organisations that serve disadvantaged patients and those that do not
- **Dependency on financial incentives**
  Relying on incentives may foster dependency on them. If provider behaviours are not ingrained, they may decline or disappear when the incentives end or new incentives are introduced
- **Demoralisation**
  Financial incentives may cause feelings of injustice and demoralisation in instances where, for example, short-term professionals receive more financial incentives than those who have established long-term practices, or where favouritism is perceived
- **Bureaucratisation**
  Results-based financing schemes may have substantial administrative costs associated with monitoring performance and managing disbursement of the financial incentives

References

1. Oxman A, Lavis JN, Fretheim A, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 1. What is evidence-informed policymaking**. Health Res Policy Syst, In Press

2. Lavis JN. **SUPPORT Tools for evidence-informed health policymaking (STP). 5. Finding systematic reviews**. Health Res Policy Syst, In Press

3. Lewin S, Oxman A, Lavis JN, Fretheim A. **SUPPORT Tools for evidence-informed health policymaking (STP). 7. Assessing the reliability of systematic reviews**. Health Res Policy Syst, In Press

4. Chopra M, Munro S, Lavis JN, Vist G, Bennett S: **Effects of policy options for human resources for health: an analysis of systematic reviews.** *Lancet* 2008, **371:** 668-674.

5. 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* 2006, **4:** 20.

6. Alderson P, Chalmers I: **Survey of claims of no effect in abstracts of Cochrane reviews.** *BMJ* 2003, **326:** 475.

7. US Surgeon General. **Surgeon General's advisory on the use of salicylates and Reye syndrome.** *MMWR Morb Mortal Wkly Rep* 1982, **31:** 289-290.

8. Hahn S, Kim S, Garner P: **Reduced osmolarity oral rehydration solution for treating dehydration caused by acute diarrhoea in children.** *Cochrane Database Syst Rev* 2002, CD002847.

9. Duley L, Henderson-Smart D: **Magnesium sulphate versus diazepam for eclampsia.** *Cochrane Database Syst Rev* 2003, CD000127.

10. Duley L, Henderson-Smart D: **Magnesium sulphate versus phenytoin for eclampsia.** *Cochrane Database Syst Rev* 2003, CD000128.

11. Villar J, Carroli G, Khan-Neelofur D, Piaggio G, Gulmezoglu M: **Patterns of routine antenatal care for low-risk pregnancy.** *Cochrane Database Syst Rev* 2001, CD000934.

12. Alderson P, Roberts I: **Corticosteroids for acute traumatic brain injury.** *Cochrane Database Syst Rev* 2005, CD000196.

13. Liberati A, Moja L, Moschetti I, Gensini GF, Gusinu R: **Human albumin solution for resuscitation and volume expansion in critically ill patients.** *Intern Emerg Med* 2006, **1:** 243-245.

14. Farquhar C, Marjoribanks J, Lethaby A, Suckling JA, Lamberts Q: **Long term hormone therapy for perimenopausal and postmenopausal women.** *Cochrane Database Syst Rev* 2009, CD004143.

15. Enayati AA, Hemingway J, Garner P: **Electronic mosquito repellents for preventing mosquito bites and malaria infection.** *Cochrane Database Syst Rev* 2007, CD005434.

16. Guyatt GH, DiCenso A, Farewell V, Willan A, Griffith L: **Randomized trials versus observational studies in adolescent pregnancy prevention.** *J Clin Epidemiol* 2000, **53:** 167-174.

17. Volmink J, Garner P: **Directly observed therapy for treating tuberculosis.** *Cochrane Database Syst Rev* 2007, CD003343.

18. Austvoll-Dahlgren A, Aaserud M, Vist G, Ramsay C, Oxman AD, Sturm H *et al*.: **Pharmaceutical policies: effects of cap and co-payment on rational drug use.** *Cochrane Database Syst Rev* 2008, CD007017.

19. Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schunemann HJ, Haines T *et al*.: **A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals.** *CMAJ* 2002, **166:** 1399-1406.

20. Grobler LA, Marais BJ, Mabunda S, Marindi P, Reuter H, Volmink J. **Interventions for increasing the proportion of health professionals practising in underserved communities**. Cochrane Database of Systematic Reviews, In Press

21. Oxman A, Fretheim A. An overview of research on the effects of results-based financing. Report Nr 16-2008. 2008. Oslo, Nasjonalt kunnskapssenter for helsetjenesten.

22. Lagarde M, Palmer N. **The impact of contracting out on access to health services in low and middle income countries**. Cochrane Library, In Press

23. 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:** 453-460.

24. Moynihan R, Oxman A, Lavis JN, Paulsen E. Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. Rapport nr. 1-2008. 2008. Oslo, Nasjonalt kunnskapssenter for helsetjenesten.

25. Fretheim A, Oxman A, Lavis JN, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 15. Monitoring and evaluation**. Health Res Policy Syst, In Press

26. UK Cabinet Office. Modernising government: presented to parliament by the prime minister and the minister for the Cabinet Office by command of Her Majesty. Cm 4310. 1999. London, Stationary Office.

27. Macintyre S, Chalmers I, Horton R, Smith R: **Using evidence to inform health policy: case study.** *BMJ* 2001, **322:** 222-225.

28. Chalmers I: **If evidence-informed policy works in practice, does it matter if it doesn't work in theory?** *Evidence & Policy* 2005, **1:** 227-42.