DRAFT: 2009 06 12

SUPPORT Tools for evidence-informed health Policymaking (STP)

11. Using balance sheets

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 11 in a series of 21 articles on tools for evidence-informed health policymaking, addresses the use of balance sheets. A balance sheet can help decision makers to develop a more accurate understanding of the important consequences of the policy options they are considering.

Objectives: In this article we identify four key questions that can be used to guide the use of balance sheets in policy decision making, and a fifth question that addresses the potential value of supplementing a balance sheet with formal economic modelling.

Key messages:
Five key questions can guide the use of a balance sheet in policy decision making to ensure that where scarce resources are used for full economic analyses, such studies focus on areas most in need of attention. These questions are:
What are the policy options that are being compared?
1. What are the most important potential outcomes of the policy options being compared?
2. What is the best estimate of the impact of the options being compared for each important outcome?
3. How confident can those applying the balance sheet be in the estimated impacts?
4. Is a formal economic model likely to facilitate decision making?

- The policy options represented in a balance sheet should be similar to those of interest to policymakers
- The outcomes that are included in the balance sheet should be those that are important to the people who will be affected. Policymakers should be cautious about using surrogate outcomes
- Estimates of the impact of policy options (compared to the status quo or another policy option) for each important outcome – and confidence intervals around those estimates – should ideally be derived from a systematic review of the best available evidence
- Typically absolute effects that are appropriate to the population(s) of interest should be used. Policymakers should be cautious about being misled by relative effects
- Judgements about the quality of the evidence for each estimate (in other words, how confident it is possible to be in terms of the estimated impact of a policy option) should take the following factors into consideration: whether the evaluations were randomised or not; whether there are possible study limitations (and therefore a risk of bias); the precision of the estimate (the width of the confidence interval); the consistency of the estimates across different studies; how directly relevant the evidence was to the populations, comparisons and outcomes of interest; and the risk of publication bias (studies with statistically significant effects may have a higher chance of being published and this could lead to an overestimation of effects)
- Formal economic models can help to make the assumptions made when comparing policy options more explicit and help to explore the effects of uncertainties and varying assumptions on the results. However, it is not always possible to develop formal economic models
- Formal economic models are likely to be most helpful in the following instances: when there is a large difference in resources consumed by the options being compared or if large capital investments, such as the building of new facilities, are required; when there is uncertainty about whether the net benefits are worth the incremental costs; and in instances where there is good quality evidence about resource consumption available

Background

This article is number 11 in a series of 21 articles on tools for evidence-informed health policymaking [1]. It is also the fifth of six articles in this series about characterising the costs and consequences of potential policy and programme options. In this article we present four key questions that can be used to guide the use of balance sheets in policy decision making, and a fifth question that addresses the potential value of supplementing a balance sheet with formal economic modelling.

A balance sheet is a simple but powerful way to present the advantages and disadvantages of different options, including policy options [2, 3]. The aim of a balance sheet is to help decision makers develop an accurate understanding of the important consequences of the options they are comparing. Balance sheets can accomplish this in a number of ways. Firstly, they condense the most important information and this allows efficient processing. Secondly, by focusing attention on the most important outcomes, balance sheets increase the likelihood that decision makers will gain an accurate perception of what is known about the impacts of the options being considered on those consequences. Thirdly, the act of constructing a balance sheet is a helpful mechanism for organising thinking, for structuring the analysis of evidence, and focusing debate. Fourthly, it can also help to develop more explicit judgements about what the most important consequences of policy options are, the underlying evidence, and subsequent judgements about the balance between the relative advantages and disadvantages of the various options. Lastly, they provide other decision makers with 'raw information', thereby allowing them to apply their own judgements about the trade-offs between desirable and undesirable consequences.

However, there are at least two limitations to using balance sheets in policy decision making. Firstly, when there are complicated trade-offs between multiple outcomes, judgements may require a high level of information processing by policymakers. Secondly, when weighing up different outcomes, the value judgements employed by policymakers could remain implicit. Formal economic modelling may help to address these limitations by making any underlying assumptions (including value judgements) more explicit, and enable the use of sensitivity analyses to explore the effects of uncertainties and varying assumptions on the results.

In describing the four key steps necessary for preparing and understanding a balance sheet (reflected in the first four questions that can be considered), and focusing on how to decide whether a formal economic model would facilitate decision making, this article draws heavily on the work of the GRADE Working Group (http://www.gradeworkinggroup.org) (see Box 7 for further information related to the GRADE assessment system) [4]. Although this group's focus has been primarily on clinical practice guidelines, these principles that relate to decisions about clinical interventions have also been shown to apply to public health and health systems decisions making [5].

Questions to consider

The following questions can be used to guide the application of a balance sheet in policy making and decisions about the value of a formal economic model to inform policy decisions:

1. What are the policy options that are being compared?
2. What are the most important potential outcomes of the policy options being compared?
3. What is the best estimate of the impact of the options being compared for each important outcome?
4. How confident can those applying the balance sheet be in the estimated impacts?
5. Is a formal economic model likely to facilitate decision making?

The first four questions are intended to guide the use of balance sheets in policy decision making. Answering the last question can help to ensure that scarce resources for undertaking full economic analyses are used where they are most needed.

Ideally, balance sheets (and economic models) should be constructed by researchers or technical support staff together with policymakers. They should also be based on systematic reviews for the same reasons described elsewhere for systematic reviews in general [6]. We will not consider the many detailed judgements that must be made when constructing a balance sheet as these have been addressed elsewhere in detail [7]. Policymakers are rarely, if ever, in a position where they are required to make all such detailed judgements themselves. However, even in instances where there is competent technical support to prepare a balance sheet, it is important that policymakers know what to look for and what questions to ask. This ensures that balance sheets can be judiciously used to inform decisions for which policy makers are accountable.

1. What are the policy options that are being compared?

When developing a balance sheet (see Table 1 for example) the first consideration is the need to identify what options are being compared. Often this is not as straightforward as it sounds. This is because those preparing a balance sheet must decide how broadly to describe both the policy option being considered *and* the comparative option. Typically, the comparison is the status quo. However, the status quo also needs to be described in relation to the new policy option being considered. Decisions need to be made, therefore, about which characteristics of the status quo are:

- Crucial – such that research with a comparison without those same characteristics would be excluded
- Important but not crucial – such that research with a comparison without those same characteristics would be included, but with less confidence that the results would be the same in the chosen setting, and
- Unimportant – such that we would be confident that the results are likely be the same in the chosen setting

These same judgements also need to be made about new policy options: which of their characteristics are crucial, important or unimportant in terms of affecting the likely impacts. For characteristics that are crucial (or when more than two policy options are being considered), it is generally desirable to prepare separate balance sheets comparing each of the options considered to the status quo.

2. What are the most important potential outcomes of the policy options being compared?

Policymakers, in general, are motivated by the desire to serve the people they represent and, as such, should be interested primarily in the impacts of policies on outcomes that are important to those affected by such policies. These include health outcomes, access to – or utilisation of – health services, unintended effects (harms), and resource use (costs or savings). Other consequences that are often important include the distribution of benefits and costs (who benefits and who pays), equity (the extent to which policy options have different (and unfair) impacts on disadvantaged and well-off populations), political acceptability and public acceptability. Ethical consequences may also be important, such as those related to a reduction in people's autonomy.

Being explicit about which outcomes are important can help to ensure that important consequences of a policy are not overlooked. It can also help to ensure that unimportant consequences are not given undue weight. This is particularly important for surrogate outcomes – i.e. outcomes that are not important in and of themselves, but are nevertheless considered important given that they are believed to reflect important outcomes. For example, people do not typically regard their blood pressure as an important concern. But what makes the issue of blood pressure important is its association with strokes, heart attacks and death, all of which *are* very much of importance to people. So when considering policies targeted at hypertension (or other cardiovascular risk factors), for example, decisions need to be based primarily on their impacts on important outcomes (cardiovascular disease), recognising that evidence of impacts on blood pressure alone is only a form of indirect evidence of the impacts on cardiovascular disease.

3. What is the best estimate of the impact of the options being compared on each important outcome?

Deciding whether the desirable impacts of a policy are worth the undesirable impacts, requires an estimate of how large the different impacts, including economic consequences, will be. Ideally this should take the form of a comparison between what could be expected for every important outcome if a policy option were to be implemented, and what could be expected if it were not – or what could be expected if an alternative policy were implemented (Box 3). In addition, it is useful to know how precise each estimate is – i.e. what the 'confidence interval' is for each estimate (see Box 4).

It is important that decision makers recognise the difference between estimates of effect that are presented as *relative* effects, and those that are presented as *absolute* effects. Patients, doctors and people making decisions about health policies and programmes, for instance, are more likely to decide to use an intervention if its effects are reported as a relative effect than if they are reported as an absolute effect [8]. For example, a study reported that 61% of the health staff population in Australia agreed to implement a colorectal cancer screening programme that would reduce the rate of deaths from bowel cancer by 17% (the relative risk reduction). In comparison, only 24% of staff agreed to implement a programme that produced an absolute reduction in deaths from bowel cancer of 0.4% (the absolute risk reduction) [9]. Both estimates were, in fact, from the same programme (for further details see Box 5).

4. How confident are we in the estimated impacts?

Six factors can lower our confidence in estimates of the impacts of a policy or programme [10]:

- A weak study design
- Other study limitations
- Imprecision
- Inconsistent results
- Indirectness of the evidence
- Publication bias

Studies in which a programme is assigned at random reduce the risk of unknown or unmeasured differences between the groups being compared, giving greater confidence that any impacts are attributable to the programme and not some other factor [11-13]. Study designs that do not use random assignment can account only for differences that are measured. For example, a study in which communities are randomly assigned to a programme or policy (such as the licensing of tobacco retailers), would provide more compelling evidence of the impacts of the policy than a study would if it compared communities that had decided themselves whether to implement a particular policy. This is because communities that decide to implement the policy are likely to differ from those that do not, in ways that could have an impact on the outcomes of interest (in this case smoking prevalence). Thus it would be impossible to know whether differences in outcomes were due to the policy or due to those other differences between the communities.

Other study limitations can affect both randomised and non-randomised impact evaluations. For example, incomplete data or the unreliable measurement of outcomes may increase the risk of an estimate being biased, and therefore lower confidence in the derived estimates.

Imprecision (in instances, for example, where there is a wide confidence interval) necessarily also lowers the confidence of ruling out chance as a factor shaping any observed differences in outcomes between compared groups (see Box 4 for an example).

If different studies of the same programme or policy have inconsistent results and there is no compelling explanation for such differences, there will also be less confidence in knowing the expected impacts arising from policy implementation.

There are several ways in which studies might not be directly relevant to a particular question, and therefore result in less confidence in the results. As noted above, if an indirectly relevant outcome (such as blood pressure) is measured in place of an important outcome (cardiovascular disease), there will be less confidence about the impacts on the important outcome (for which the indirect outcome is a surrogate). If only indirect comparisons are provided, confidence will be lower. Similarly, if studies of an implementation strategy compared to a control (with no intervention) and other studies of a different implementation strategy compared to a control are provided, but there are no head-to-head comparisons of the two implementation strategies, we would be less confident in the results than if a head-to-head comparison had been provided. Other ways in which evidence can be indirect include differences between a study and the setting of interest in the characteristics of the population, the policy or the comparison that could affect the magnitude of the impact.

Studies that find statistically significant effects are often more likely to be published than ones that do not find statistically significant effects [14]. When such 'publication bias' appears likely, confidence in estimates from published studies alone may also be lowered. Publication bias should be considered in instances where, for example, there are a number of small studies, especially if those small studies are industry-sponsored, or if the investigators are known to share other similar conflicts of interest.

In summary, assessments of the 'quality' or robustness of evidence, and confidence in estimates of the likely impacts of policy options, depend on a consideration of *all* of the factors noted above. Although there are no fixed rules for assessing these factors, judgements about the quality of evidence that explicitly address each factor reduce the likelihood of overlooking important factors and reduce the probability of bias (see also Box 6 for an example). Using a systematic and transparent approach, such as the GRADE approach (see Box 7), make it easier to inspect the judgements that were made [4].

5. Is a formal economic model likely to facilitate decision-making?

Formal economic models, such as cost-effectiveness analyses and cost-utility analyses, can help to inform judgements about the balance between the desirable and undesirable consequences of a policy option [3]. Unfortunately, published cost-effectiveness analyses, particularly those undertaken for drugs, have a high probability of being flawed or biased, and are not specific to a particular setting [15]. Policymakers may therefore consider developing their own formal economic models. However, to do this they must have the necessary expertise and resources.

Economic models can be valuable for complex decision making and for testing how sensitive a decision is to key estimates or assumptions. However, a model is only as good as the data on which it is based. When estimates of benefits, harms or resource use come from low quality evidence, the results will necessarily be highly speculative (see Box 8 for an example).

A full economic model is more likely to help to inform a decision when there is:

- A large difference in the resources consumed between the compared options
- Large capital investments are required, such as the construction of new facilities
- Uncertainty about whether the net benefits are worth the incremental costs
- Good quality evidence regarding resource consumption

An economic model can also be used to clarify information needs by exploring the sensitivity of an analysis to a range of plausible estimates.

Resources

Useful documents and further reading

- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ, and the GRADE Working Group. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924-6.

- Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, and the GRADE Working Group. What is 'quality of evidence' and why is it important to clinicians? BMJ 2008; 336:995-8.

- Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Vist GE, Schunemann HJ, and the GRADE Working Group. Incorporating considerations of resource use. BMJ 2008; 336:1170-3.

Links to websites

- GRADE Working Group http://www.gradeworkinggroup.org/

Box 1. What is being compared? Case example: The licensing of tobacco retailers

Reducing teenage smoking was a priority for the Minister of Health in a European country. A report of policy options for achieving this was commissioned by the government concerned and a report was prepared by leading public health experts. One of the policy options considered in the report was the licensing of tobacco retailers, with the loss of such a license as a penalty for the illegal selling of tobacco to minors. This option was compared to the status quo, namely the absence of licensing for tobacco retailers. The public health experts did not undertake or use a systematic review, nor did they specify which characteristics of the policy option (or comparator) they considered to be crucial or important.

Important differences between the status quo of the areas where the policy makers considered implementing the policy and those where the studies were done, might have been included. Such considerations may have included, for example, other policies already in place to reduce the sales of tobacco to minors. For example, existing legislation may already have made the sale of tobacco to minors illegal, or contained other methods by which the legislation could be enforced (e.g. fines or other penalties for illegal sale of tobacco, face-to-face education of retailers (informing them about the legal requirements), or media campaigns to raise community awareness). There might also have been differences in the ease with which minors could obtain tobacco from other sources (e.g. from parents, friends or through theft).

The experts explicitly considered two policy options for the licensing of tobacco retailers: three compliance checks per year (by a teenager attempting to purchase tobacco) to make sure that retailers were not selling tobacco to minors, and one compliance check per year together with internal control (requiring retailers themselves to control that tobacco is not being sold to minors). The penalty for non-compliance in both cases was the loss of the relevant license. Although other additional ways of enforcing the licensing can be imagined, and some of which have already been evaluated, the experts did not explicitly address whether these differences were likely to result in important differences in the effectiveness of the policy.

Box 2. What are the most important outcomes? Case example: The licensing of tobacco retailers

The primary outcome considered by the expert report commissioned by the government concerned was the prevalence of smoking. This was recognised to be a surrogate outcome for the consequences of smoking. The impact on life years saved was estimated based on the estimated impact on the prevalence of smoking and on epidemiologic data linking smoking to mortality. Impacts on morbidity were not considered. Other impacts that were explicitly considered were administrative costs, political acceptability and public acceptability. Costs to retailers and potential harms (e.g. increased theft or cross-border shopping) were not addressed. The report also did not address who would pay the administrative costs of such schemes or the potential differences in the impacts of the policy on different populations (e.g. socio-economically disadvantaged minors or those living close to the border (who could potentially cross the boarder to purchase tobacco). It also did not address the ethical consequences (e.g. those related to using a minor or someone pretending to be a minor for compliance checks, or the fairness of the policy in relation to the potentially different impacts on different groups of minors and different retailers).

Box 3. What are the best estimates of the impacts? Case example: The licensing of tobacco retailers

The expert report on policies to reduce teenage smoking commissioned by the government concerned estimated that licensing tobacco retailers would result in a 10% relative reduction in the number of smokers. Given the current prevalence of smokers, the absolute effect was estimated to be 1,650 fewer smokers per year. Based on epidemiological models of the increased risk of dying due to smoking, the experts estimated that this policy would save 9,240 lives per year. No confidence intervals were provided, although it was noted that the actual effect was very uncertain and a range of estimates was used to calculate the cost-effectiveness of licensing tobacco retailers. Administrative costs were estimated, based on an estimate of how many retailers sold tobacco, an assumption about what it would cost to process each license, and an assumption about what each inspection (to check compliance with the requirement to not sell tobacco to minors) would cost. Using these different assumptions, the total cost was estimated to be between €7.2 million and €10.5 million per year. The report estimated that the political acceptability of the policy was “low” and that public acceptability was “relatively high”. The basis for these estimates was unclear.

Box 4. Confidence intervals

A confidence interval (CI) is the range around an estimate which conveys how precise the estimate is. The confidence interval is a guide that represents how sure it is possible to be about the quantity we are interested in (e.g. the effect of a policy option on an outcome of interest). The narrower the range between the upper and lower numbers of the confidence interval the more confident it is possible to be about what the true value is. The wider the range, the less certain it is possible to be. The width, or range, of the confidence interval reflects the extent to which chance may be responsible for an observed estimate (wider intervals reflect the greater likelihood of chance being a factor). A 95% CI means that we can be 95% confident that the true size of an effect is between the lower and upper confidence limit. Conversely, there is a 5% chance that the true effect is outside this range.

Box 5. Relative and absolute effects

Relative effects are ratios. For example, a risk ratio (RR) is the ratio between the risk in an intervention group and the risk in a control group. If the risk in an intervention group is 2% (i.e. 20 per 1,000) and the risk in a control group is 2.4% (i.e. 24 per 1,000), the risk ratio (or relative risk) will be 20/24 or 83%. 'Relative risk reduction' is another way of expressing relative effects. This is the proportional or percentage reduction in risk, and is equal to 1-RR which, in this case, is 17%.

If the RR value is exactly 1.0, this means that there is no difference between the occurrence of the outcome in the intervention group and the control group. But the significance of this value being above or below this 1.0 depends on whether the outcome being measured is judged to be good or bad. If the RR value is greater than 1.0, the intervention increases the risk of the outcome. If the desired outcome is considered to be good (for example, the birth of a healthy baby), an RR greater than 1.0 indicates a desirable effect for the intervention. Conversely, if the outcome is bad (for example, death) an RR value greater than 1.0 would indicate an undesirable effect. If the RR value is less than 1.0, the intervention decreases the risk of the outcome. This then indicates a desirable effect, if it is a bad outcome (for example, death) and an undesirable effect if it is a good outcome (for example, birth of a healthy baby).

Absolute effects are differences. For example, absolute risk reduction (ARR) is the difference between the risk with the intervention and the risk without the intervention. In this example the ARR is 2.0% (20 per 1,000) less 2.4% (24 per 1,000) – or 0.4% (4 per 1,000) fewer deaths from bowel cancer.

Box 6. How confident are we in the estimated impacts? Case example: The licensing of tobacco retailers

The expert report commissioned by the government concerned concluded that the empirical basis for the licensing of tobacco retailers was "robust". However, the basis for this judgement was unclear. The experts did not conduct, or cite, a systematic review as the basis for their estimates, although a systematic review was available [16]. An assessment of the evidence summarised in the systematic review using the GRADE approach, in contrast to the experts' unexplained judgement, suggests that the quality of the evidence was very low for all the important outcomes (see Box 7 for further information related to the GRADE assessment system). Table 1 summarises the findings of the expert's report in the form of a balance sheet for this policy decision and shows an assessment of the quality of the evidence for the three estimates using the GRADE approach.

The authors of the systematic review (which included a broader range of interventions and study designs) concluded: "Interventions with retailers can lead to large decreases in the number of outlets selling tobacco to youths. However, few of the communities studied in this review achieved sustained levels of high compliance. This may explain why there is limited evidence for an effect of the intervention on youth perception of ease of access to tobacco, and on smoking behaviour." The "pessimistic" estimates of benefits in Table 1 are consistent with the findings of the systematic review (and were not considered in the expert report).

Box 7. The GRADE system for assessing the quality of evidence

Evaluating the quality of the evidence is a judgement about the extent to which one can be confident that an estimate of effect is correct. GRADE provides a systematic and transparent approach to making these judgements for each outcome that is important to a decision [10]. The judgements are based on the type of study design (randomised trials versus observational studies), the risk of bias (study limitations), the consistency of the results across studies, and the precision of the overall estimate across studies. Based on these considerations for each outcome, the quality of the evidence is rated as high, moderate, low or very low using the following definitions:

⊕⊕⊕⊕ **High**	Confident that the true effect lies close to that of the estimate of the effect
⊕⊕⊕◯ **Moderate**	The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
⊕⊕◯◯ **Low**	The true effect may be substantially different from the estimate of the effect
⊕◯◯◯ **Very low**	Very uncertain about the estimate

Box 8. Is a formal economic model likely to help? Case example: The licensing of tobacco retailers

The expert report commissioned by the government included an economic analysis. This concluded that the cost per life year saved by licensing tobacco retailers and conducting compliance checks, was between approximately €900 and €92,000, with a best estimate of €8,000. The authors noted that there was substantial uncertainty about their estimates and suggested focusing on the range of estimates rather than the best estimate. Nevertheless, they reported exact estimates (based on the assumptions they made) and concluded that the empirical basis for recommending licensing tobacco retailers was robust. As a result, policymakers who failed to read this report critically could conclude (and in our opinion, wrongly) that the report provided high quality evidence that the licensing of tobacco retailers was as cost-effective as (or more cost-effective than) a wide range of clinical preventive services paid for by the government concerned. A more systematic review of the underlying evidence [16], and a summary of the findings that included more systematic and transparent judgements of the quality of the evidence (as shown in Table 1), would have provided a better basis for decision making.

Table 1. Should tobacco retailers be licensed conditional on their not selling tobacco to minors?

Population: Minors (as defined by a legal age limit)
Setting: Europe
Interventions: Licensing of tobacco retailers + compliance checks [*]
Comparison: No licensing or compliance checks

Outcomes	Impact			Number of studies	Quality of the Evidence (GRADE) [†]
	Pessimistic	Best guess	Optimistic		
Reduced number of smokers per year	0	?	1,650 in the country (population 4.5 million)	4	⊕◯◯◯ Very low [‡]
Life years saved per year	0	?	9,240 in the country (population 4.5 million)	4	⊕◯◯◯ Very low [§]
Cost per year	€10,5 million (3 controls per year)	?	€7,2 million (1 control per year + internal control)	0	⊕◯◯◯ Very low [**]

[*] The proposed licensing law in the European country in question would require retailers to have a licence to sell tobacco. The policy options that were considered included three compliance checks per year, and one per year together with internal control. Compliance checks (by a teenager attempting to purchase tobacco) are done to ensure that tobacco is not being sold to minors. The penalty for non-compliance is loss of the licence. Internal control requires the retailers themselves to have routines for controlling that tobacco is not being sold to minors.

[†] See Box 7.

[‡] The systematic review used as a basis for this summary included one relevant randomised trial and three controlled before-after studies with important limitations. There was a high risk of bias for the estimated impacts on smoking prevalence. There was also important inconsistency in the results without a compelling explanation for those. The studies in the review were from the United States (2), the United Kingdom (1) and Australia (1) with differences in the interventions and uncertainty about whether similar results would be expected where this policy was being considered. Two studies found an effect in lower age groups that was not sustained in one study and two studies did not find a change in smoking behaviour. Based on these studies it is difficult to estimate what the best estimate of the impact of licensing of tobacco retailers with compliance checks would be on reducing the number of smokers. A lower estimate would be no impact. The upper estimate is taken from an expert report (see Boxes 1-3).

[§] The upper estimate of life years saved, which is taken from the same expert report, has the same limitations as the estimate of the impact on smoking behaviour, since it is based on that estimate. In addition, it is based on assumptions about what would happen long beyond the length of the studies that evaluated impacts on smoking behaviour and assumptions about the impact of the changes in smoking behaviour on mortality.

[**] The estimates of the cost of the policy are taken from the expert report. They are based on an estimate of how many retailers sold tobacco, an assumption about what it would cost to process each licence and an assumption about what each compliance check would cost.

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. Eddy DM: **Comparing benefits and harms: the balance sheet.** *JAMA* 1990, **263:** 2493, 2498, 2501.

3. Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A *et al*.: **Incorporating considerations of resources use into grading recommendations.** *BMJ* 2008, **336:** 1170-1173.

4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P *et al*.: **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.** *BMJ* 2008, **336:** 924-926.

5. **The SUPPORT Collaboration. How SUPPORT Summaries are prepared** [http://www.support-collaboration.org/summaries/methods.htm . 2009].

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

7. Guyatt G, Oxman A, Kunz R, Vist G, Brozek J, Norris S *et al*.. **GRADE guidelines: 1. Introduction – GRADE evidence profiles and Summary of Findings tables**. J Clin Epidemiol, In Press

8. Akl E, Oxman A, Herrin J, Vist G, Terrenato I, Sperati F *et al*.. **Using alternative statistical formats for presenting risks and risk reductions**. Cochrane Database of Systematic Reviews, In Press

9. Adily A, Ward J: **Evidence based practice in population health: a regional survey to inform workforce development and organisational change.** *J Epidemiol Community Health* 2004, **58:** 455-460.

10. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ: **What is "quality of evidence" and why is it important to clinicians?** *BMJ* 2008, **336:** 995-998.

11. Kleijnen J, Gotzsche P, Kunz R, Oxman A, Chalmers I: **So what's so special about randomisation.** In *Non-Random Reflections on Health Care Research: On the 25th Anniversary Of Archie Cochrane's Effectiveness and Efficiency*. Edited by Chalmers I, Maynard A. London: BMJ Publishers; 1997:93-106.

12. Kunz R, Oxman AD: **The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials.** *BMJ* 1998, **317:** 1185-1190.

13. Savedoff WD, Levine R, Birdsall N. When will we ever learn? Improving lives through impact evaluation. 2006. Washington DC, Center for Global Development.

14. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: **Publication bias in clinical trials due to statistical significance or direction of trial results.** *Cochrane Database Syst Rev* 2009, MR000006.

15. Friedberg M, Saffran B, Stinson TJ, Nelson W, Bennett CL: **Evaluation of conflict of interest in economic analyses of new drugs used in oncology.** *JAMA* 1999, **282:** 1453-1457.

16. Stead LF, Lancaster T: **Interventions for preventing tobacco sales to minors.** *Cochrane Database Syst Rev* 2005, CD001497.