

7 APPRAISAL AND EVALUATION

7.1 INTRODUCTION

This chapter introduces the important topics of appraisal and evaluation of road user charging systems. Appraisal is an “ex-ante” activity, defined as work carried out in advance of implementing a policy of any kind, such as a road pricing scheme, in order to quantify and assess its expected impacts. Appraisal helps guide the decision on whether to implement the scheme and whether to make changes to the design of the road pricing scheme.

In contrast, evaluation is an “ex-post” activity, defined as work carried out after implementation to find out what the actual impacts of the policy have been. Evaluation studies add to the body of literature available to improve our ability to predict and appraise future schemes.

Both appraisal and evaluation should be based on the full list of objectives of the road pricing scheme, as discussed in Chapter 2. Ideally, evaluation should also consider also the objectives of interest to other cities, so that the results are potentially able to be transferred to other locations, as discussed in Chapter 12.

7.2 WHAT IS THE IMPORTANCE OF THE THEME?

7.2.1 What do we mean by appraisal and evaluation?

Appraisal is the process of assessing, in advance of implementing a policy such as a road pricing scheme, whether it will be effective in meeting the city’s objectives and, where appropriate, whether it will satisfactorily avoid any constraints which the city places on the scheme. At its simplest, it can be applied to one proposal, and used to answer the simple, but often challenging question: should we implement this scheme? However, with as major a change as the implementation of road pricing, it is common to develop a number of policy options and compare them to produce a shortlist of possible schemes. As an example, the London Congestion Charging study of the 1990s considered over 40 possible design options, and appraised them against a common set of objectives (May *et al*, 1996). Less commonly, but equally importantly, appraisal can be used at an early stage in the development of options to identify weaknesses in them which can be overcome. For example, if a given design is effective in reducing congestion, but adversely affects a particular group of users, they could be provided with an exemption.

Thus appraisal can be used to answer three types of question:

- Does this design option have weaknesses which could potentially be improved?
- Which of these design options performs best against our objectives?
- Should we implement the best design option?

Evaluation, by contrast, is the process of assessing, after a policy such as road pricing has been implemented, whether it was successful in meeting the objectives set for it, and avoiding any constraints. It is common for this step in the policy cycle to be overlooked. Politicians may understandably take the view that, once a scheme has been introduced and broadly accepted, it is better to move on to the next policy challenge than to spend time and resources checking in detail how well it worked. With a scheme as complex as road pricing, however, it is unlikely, as discussed in Chapter 11, that the public will readily accept all aspects of the scheme once it has been implemented. At the very least, therefore, evaluation will provide a means of obtaining the evidence with which to answer the scheme’s continuing critics and to increase public acceptance. In practice, it is probable that some aspects of the scheme will not work as well as planned. In such cases, evaluation provides the evidence that improvements are needed, and the basis for enhancing the scheme’s effectiveness.

More generally, with a policy as novel as road pricing, evaluation provides empirical evidence which will help other cities decide whether they would benefit from a similar policy. The series of annual evaluations of the London Congestion Charging Scheme (TfL, 2007) and, more recently, the evaluation of the Stockholm scheme (City of Stockholm, 2006), are already being used internationally as input to other cities' policy decisions. Finally, as discussed in Chapter 6, comparison between evaluation of the actual outcome, and appraisal of the predicted outcome, will help to assess whether the process of prediction was effective, or whether enhancements are needed to the predictive models used.

Thus evaluation can be used to answer four types of question:

- Did the scheme work effectively in meeting our objectives?
- Are there aspects of the scheme which merit improvement?
- Were the predictions accurate, or are improvements needed in the prediction process?
- Is this scheme (in another city) something which I should be considering for my city?

While the first three of these questions can focus solely on the objectives and constraints of the city which implemented the scheme, evaluation to satisfy the fourth question should ideally cover all the objectives which other cities might have.

7.2.2 How important is appraisal to decision makers?

Appraisal, considered independently in the UNAQ results, was only ranked 5 out of the 9 themes. It is difficult to determine why this was. It may be that most cities answering the questionnaire were at too early stage in considering design options to have given much thought to appraisal. However, prediction was ranked second and, as noted below, it is difficult to enter into the process of prediction without considering the types of answers which are needed, as input to the appraisal process. It may be that respondents were implicitly grouping prediction and appraisal together in their responses. In practice, the importance of appraisal to decision-makers cannot be overestimated, as the outcomes are critical to decisions about whether to implement a given scheme and are often the primary basis for choosing between alternative options.

Evaluation was not identified separately in the UNAQ. In practice, as noted above, it is likely to be critical in answering the continuing critics of a scheme. In the longer term it is extremely important, to politicians, professionals and researchers, in order to further knowledge about the predicted and observed impacts of road pricing schemes. However, in the aftermath of an implementation, there may be a danger that it will not be given the priority it deserves by local decision-makers. The history of transport planning innovations suggests a general lack of "after" studies, probably because by that stage the major political decisions have been taken and thus such studies have not been seen as having a critical role to play in future policy. In the case of road pricing, the level of academic, political, media and public interest in the relatively small number of implementations to date has encouraged the collection and dissemination of a considerable volume of "after" data. However, there remains a challenge to ensure that evaluation continues to further our understanding of the impacts of road pricing (such as through common indicators that can be compared across different schemes). The international community should therefore encourage all cities to carry out comprehensive evaluations of the schemes which they implement.

7.2.3 What should we appraise and evaluate?

The answer to this question may seem obvious: we should be appraising the set of options for road pricing, and evaluating the scheme as implemented. But it is important to define carefully the boundaries of what is appraised or evaluated. As noted in Chapter 3, a scheme can be the road pricing cordon or area itself, with the set of charges, discounts and exemptions applied. Even at this level, the appraisal or evaluation should extend to consider the performance of the technology (Chapter 4) and the administrative and enforcement processes (Chapter 5). The annual evaluation of the London scheme (TfL, 2007) provides an example of good practice in all of these.

But one of the principal purposes of a road pricing scheme may well be to generate revenue. If this is the case, should the appraisal simply assess the level of revenue generation, and the possible benefits of having access to that revenue? Or should the package be the road pricing scheme and the enhancements which that revenue has financed? In both London and Stockholm, much of the revenue has been used to finance improvements in public transport, and in both cases the evaluation has extended to assessing the (enhanced) performance of the public transport network.

7.2.4 How do appraisal and evaluation relate to the objectives of road pricing?

Looked at another way, road pricing will typically only be successful if it is implemented as part of a package of measures, which will include public transport improvements, and will probably also include traffic management measures to facilitate diversion and to enhance environmental benefits. Had the road pricing scheme been implemented (or designed) without these complementary measures, it would have performed differently. Thus there is a strong case for specifying the boundaries to include the full package of measures and appraising or evaluating that package. This makes the scale of the appraisal and evaluation task greater, but it will substantially improve the understanding of what might happen or has happened.

As noted above, appraisal should assess whether a scheme will meet the city's objectives, and evaluation should assess whether it has met them. As noted in Chapter 2 it is essential that cities enter the process of road pricing scheme design with a clear understanding of their objectives, and of the relative importance of them. In some cases, less direct objectives can be expressed as constraints; for example: "to reduce congestion and enhance the environment, subject to not adversely affecting the economy". These same objectives and constraints should be used in the appraisal and evaluation processes. Some appraisal procedures permit the objectives to be weighted to express their relative importance (see Section 7.3).

Such weighted appraisals are probably not appropriate to evaluation. Instead, the focus of evaluation should be on whether the scheme has performed well against each objective, taken on its own. A comprehensive evaluation should, as noted earlier, consider each of the objectives listed in Chapter 2, even if some of them were not critical for the city concerned.

7.2.5 How do appraisal and evaluation relate to prediction?

For most of the objectives listed in Chapter 2, it will not be self evident how a particular scheme would perform. Instead it will be necessary to carry out a formal process of prediction to estimate how travel patterns might change, and what impacts these would have on outcomes such as congestion, accidents, air pollution and the economy. Thus the prediction process informs the appraisal, and should be designed to provide the information which the appraisal process requires. In practice, therefore, it is better to specify the appraisal process first before finalising the prediction methods.

As noted in Chapter 6, outcomes for some objectives, such as impacts on the urban economy, will be harder to predict than others. There is thus a danger that they will be overlooked. An effective appraisal process should be aware of these weaknesses in prediction, and take steps to allow for uncertainty in such impacts.

The link between prediction and evaluation is very different. The evaluation process does not need prediction, since impacts have occurred and can in most cases be measured. However, a comparison between appraisal and evaluation can indicate how well the prediction process has worked. Effective evaluation is thus central to enhancing the prediction process.

7.3 WHAT APPRAISAL METHODS ARE AVAILABLE?

7.3.1 What are the principles of appraisal?

Appraisal of a policy against a single objective is relatively straightforward. Provided that the objective can be measured by an indicator, and performance against it predicted, the policy which performs best will be the one that generates the highest value of that indicator. It would be possible, for example, to appraise road pricing schemes in terms of their ability to generate revenue in this way. Taking this simple example further, it may be that all that is needed is to generate more than a certain level of revenue. This can be reflected by setting an achievement target for the indicator of revenue. All schemes which met that target would be acceptable. Targets may be particularly appropriate for reflecting constraints.

In practice, policies will be appraised against several objectives, which makes the process much more complex. Firstly, different objectives will be measured in different ways. Revenues can be measured in money terms, but it is not clear that pollutants or accidents can, unless money values can be assigned to them. Some objectives, such as liveability, are not even readily quantified. The equity objective, in particular, needs to be measured in terms of a range of values, and any indication of distribution of costs and benefits will be lost if too aggregate a measure is used. Thus objectives differ substantially in the extent to which they can be measured or be assigned money values.

Moreover, if some objectives are more important than others, it may be appropriate to assign weights to them, so that a given value of reduction in congestion can be reflected as being, say, twice as important as a given value of the reduction in pollution. Such a process requires the weights to be specified, perhaps by politicians, stakeholders or the public, and this in turn requires the objectives to be measured in a comparable way.

The appraisal process, and the predicted information which is input to it, will be somewhat uncertain. Uncertainties can be found both in the input data and in the weights assigned to different objectives. The issue of optimism bias, as discussed in Chapter 6, is an important example of uncertainties in input data. Ideally an appraisal process should reflect these sources of uncertainty, and enable the user to assess how robust the resulting decision is.

Finally, the timescale over which costs and benefits arise will affect the appraisal. In general it is assumed that a given benefit obtained in year one will be more valuable than the same benefit achieved in year two or year ten. However, some benefits, such as those to future generations, will only be of relevance in a future year, and cannot be treated in the same way.

Thus ideally any appraisal method needs to reflect:

- 1 the use of targets for some or all objectives, and particularly for constraints;
- 2 the differing degrees to which different objectives can be measured;
- 3 the differing degrees to which different objectives can be valued;
- 4 the assignment of different weights to different objectives;
- 5 the uncertainties which arise both in prediction and in the assignment of weights;
- 6 the relative value of costs and benefits arising in different years.

7.3.2 What are the principal appraisal methods?

There is an extensive literature on appraisal methods (e.g.: Odgaard *et al*, 2005; Grant-Muller *et al*, 2001; Mackie *et al* 2001). We provide only a brief discussion of the principal methods. The principal distinction is between Cost Benefit Analysis (CBA) and Multi-Criteria Appraisal (MCA).

Cost Benefit Analysis makes the simplifying assumptions that performance against all objectives can be quantified, and that all those quantified impacts can be assigned money values. It has been common practice for some time to assign money values to time and

accident savings, although there remain concerns about, for example, the treatment of small time savings and the ethics of attempting to value life. More recently some countries have assigned money values to air pollutants, noise and contributors to global warming. However, it remains difficult to value performance against objectives such as liveability, health and economic growth. In particular aggregation of all benefits in money terms makes it difficult to demonstrate the distributional (or equity) impacts of a policy.

Cost Benefit Analysis also traditionally specifies a time horizon over which costs and benefits are predicted, which may range from 20 years to as much as 60 years in current UK practice (DfT, 2008). Standard discount rates are applied to each year's impacts, so that they may be aggregated as a net present value of costs or benefits. Test discount rates are usually set nationally, and typically range from 2% to 8%. This approach makes it difficult to place any emphasis on the benefits to future generations, since they will occur in a future year, and be assigned a small value once discounted.

Traditional Cost Benefit Analysis makes little attempt to reflect uncertainties in the input data, though recent work has introduced techniques for dealing with optimism bias (DfT, 2008). Weights for different objectives are implicit in the money values, and it is unusual to test sensitivity to different assumptions on money values or discount rates.

Thus CBA makes no attempt to use targets ((1) above); makes simplifying assumptions on (2), (3) and (4) above; only considers uncertainty in input values ((5) above); but does adopt a process for dealing with costs and benefits over time ((6) above).

Multi-Criteria Appraisal, by contrast, avoids the pitfalls of assigning money values and instead uses one or more indicators to measure performance against each objective. Where the objective is readily quantified, as for example with reduction of accidents or air pollution, these quantified indicators are used directly. Even so, there is an extensive debate on the appropriateness of different indicators (Marsden *et al*, 2006). Where objectives are more qualitative, it is common practice to use semantic scaling to convert them into quantified indicators (Jopson *et al*, 2007; Kelly *et al*, 2008).

It is then possible to set targets for each of these indicators, and to appraise a policy in terms of its ability to achieve this set of targets. Whether or not this is done, the performance of a policy can be presented separately, usually in tabular form, against each indicator. The appraiser then has to compare the sets of results across objectives and across policy options to decide which options are the most acceptable. This can be a challenging task if the data is extensive, but this approach has the advantage of leaving the responsibility for deciding on the relative importance of different impacts to the decision-maker.

As an alternative, it is common practice to assign weights to different objectives, or to their related indicators, and to generate a weighted sum of the performance against all objectives. In a similar vein, weights could be assigned to results occurring in different years. Such methods can produce a single performance indicator, much as CBA does, but with a more explicit process of valuation. This makes the selection process easier, but may disguise some of the assumptions involved.

Several MCA methods incorporate sensitivity tests which allow the effect of differing input values, valuation methods and weights to be tested, either to produce different ranked lists of options or to assess the robustness of a preferred scheme to different assumptions. Such methods offer a much more robust approach to appraisal, but can be more complex in use.

Thus MCA methods are able to reflect the use of targets ((1) above), though this is rarely done; they can address differing abilities to quantify and value different impacts ((2) and (3) above); and they have well developed approaches for dealing with weighting and uncertainty ((4) and (5) above); but they are typically less effective in dealing with the distribution of impacts over time ((6) above).

Finally, some established methods adopt a combination of CBA and MCA approaches. This is particularly the case with the UK's NATA method (DfT, 2008). One of the latest studies of

appraisal methods in the EU states (excluding Luxemburg) and Switzerland is from the HEATCO project (Odgaard *et al* 2005). It concluded that the standardisation of principles for project appraisal varies considerably across countries and modes. In addition, differences also can be found in the way CBA approaches are employed. In most of the countries, CBA is used as a means to choose between different project alternatives (including “doing nothing”), to demonstrate the need for a measure and/or to prioritise between different variants. These findings support the earlier work of Grant-Muller *et al* (2001) who in addition stated that “*All appraisal frameworks contain a mixture of monetized impacts, impacts measured in both physical and qualitative forms*”.

Obviously variations in appraisal approaches/assumptions, even within the EU and Switzerland, can affect the transferability of the model predictions from, an issue taken up in Chapter 12 of this report.

7.3.3 How might the appraisal of road pricing schemes differ from appraisal of other policy instruments?

Recent draft guidance on the appraisal of road pricing schemes in the UK (DfT, 2008) suggests that in many ways the appraisal of road pricing can follow the principles applied to the appraisal of any policy instrument. However, it highlights five ways in which road pricing appraisal raises specific issues.

The first issue relates to the complexity of the responses to a road pricing scheme, which in turn affect the performance in terms of efficiency. This is particularly a challenge for prediction, but the extent of transfers between modes and times of day, and the potential for suppression of journeys mean that the calculation of benefits and disbenefits is a complex process.

The second issue concerns the wider economic benefits, which are addressed in more detail in Chapter 8. The processes by which changes in travel costs and in accessibility affect business and residential location are poorly understood, and the extent to which benefits and losses to businesses exceed those calculated directly from changes in travel costs are uncertain. This issue relates to the treatment of agglomeration and productivity benefits, which is considered in more detail in Chapter 8. Once again, it is a particular challenge for road pricing schemes because of the extent of the influences, both positive and negative, on business costs.

The third issue relates to the treatment of changes in accessibility. The DfT guidance raises the specific issue of changes in severance, but in practice this will be only one of a large number of changes in accessibility for the users of all modes. These will in turn be one of the principal factors affecting the distribution of benefits.

The distribution of benefits and the equity implications are the fourth issue raised. The DfT guidance recommends an approach to the modelling and appraisal of equity implications which has proved extremely challenging for the cities involved.

Finally the guidance notes that the treatment of value for money (as reflected in UK practice by the comparison of CBA net present benefits and MCA outcomes with net present costs) needs to be modified to reflect the fact that road pricing will usually generate surplus revenue. Despite this observation, guidance on the ways in which the value of revenues generated, for example using shadow pricing [May *et al*, 2000, May *et al*, 2001] is still limited.

7.3.4 What can we learn from past approaches to appraisal and evaluation?

Appendix B includes several examples of approaches to the appraisal of road pricing proposals and the aggregate evaluation of implemented schemes. It demonstrates that most appraisals and evaluations have attempted, at least, to reflect the full range of policy objectives, and that they have typically used a combination of CBA and MCA approaches.

Little attempt has been made to weight the different objectives, and the treatment of distributional and equity impacts remains weak.

A particular challenge to appraisal and evaluation is raised by the two studies by Prud'homme *et al*, and their comparison with evaluations of London and Stockholm conducted by others. Both attempt an aggregate Cost Benefit Analysis, which inevitably disguises the assumptions made as to values assigned. Part of the difference in overall assessments relates to differences in the values assigned to travel time and accidents. However, more subtly Prud'homme *et al* adopt a different approach to estimating the travel time changes outside the charging zone, include the costs, but not necessarily the benefits, of provision of new buses, estimate the costs of additional overcrowding on rail, but do not assess the benefits of greater reliability on the road network. These two examples demonstrate the importance of having an agreed definition of the scope of the appraisal, in terms of objectives covered, policy instruments included and user impacts allowed for. It also highlights the danger of simply aggregating a set of disparate impacts, and the need, if this is to be done, to agree on the values to be assigned to each type of impact.

7.4 WHAT IMPLICATIONS DOES APPRAISAL HAVE FOR OTHER THEMES?

There are the following implications for other themes:

- **Objectives:** Fundamentally, the appraisal process provides indications as to the achievement of the overarching objectives of the road user charging scheme. It is essential, therefore, that these objectives are clear at the outset.
- **Scheme Design:** The appraisal outcome is dependent on the scheme design and analysis of the impacts of the scheme can provide insight into ways to iteratively refine the scheme so as to augment the positive and mitigate the negative impacts of the scheme. In addition, appraisal provides information for complementary measures that may be required in support of a package of measures for which road pricing is an instrument. For example, if the appraisal indicates a possible increase in public transport patronage and significant mode switching, adequate capacity could be provided on the public transport capacity (e.g. by running additional buses as in the case of London).
- **Technology and Business Systems :** Part of the appraisal process should be the verification that the technology and Business Systems are compatible with the operations of the road user charging system. The London experience suggested that if costs could be reduced, then the benefits would be significantly increased.
- **Prediction:** To a large extent, the prediction and appraisal processes are usually conducted simultaneously. We draw a reasonably artificial distinction in this report for expository convenience. Clearly information from prediction methodologies is essential to the appraisal process. Indeed the prediction method should ideally only be finalised once the appraisal needs are known. Thus these two activities are inextricably linked.
- **Economy:** Analysing the economic implications of the scheme should form part of the appraisal process. In addition, the appraisal process can also provide clues as to the business groups that might be affected by the pricing scheme design.
- **Environment:** Environmental impacts are an essential element in the appraisal process. However, some environmental impacts are difficult to quantify, and for others the second order implications on humans and ecology are difficult to determine.
- **Equity:** One key element of appraisal is to assess the equity aspect of the scheme. A positive Net Present Value or Economic Benefit does not automatically imply that a scheme will be equitable. The information from appraisal can be used to gauge the vertical and horizontal impact groups of the schemes. Ways to reduce real and perceived inequity can then subsequently be sought again probably through an iterative refinement process.
- **Acceptability:** It is important to point out here that a scheme that ticks all the right boxes in the appraisal may still not be deemed acceptable by the public. However if

careful analysis has been undertaken to identify the scheme impacts of various socio-economic groups, then steps can be taken to increase acceptability.

- **Transferability:** Scheme appraisal methods differ even amongst EU countries. It is thus difficult to ensure that true comparisons are being made. From this standpoint, an MCA appraisal which does not attempt to weight or aggregate impacts may better facilitate the transfer of results between cities.

7.5 WHAT ARE THE GAPS IN OUR KNOWLEDGE?

The limitations in our understanding of appraisal methods can be considered in two groups: those which relate to the specific requirements for the appraisal of road pricing schemes (Section 7.3.3) and those which relate more generally to gaps in appraisal methodology, but which are relevant to road pricing.

7.5.1 The specific requirements for appraising road pricing schemes

Section 7.3.3 identified five ways in which appraisal of road pricing differs from that of other policy instruments.

The first related to the complexity of responses. The prediction process needs to estimate the scale of change in journey frequency, destination, timing, mode and route, and the second order impacts on the transport network. Some of the changes, for example in flows and modal shares, will be substantial, and appraisal methods designed to reflect marginal changes will not be appropriate. Guidance is needed on the assumptions which it is appropriate to make, and their implications for the appraisal process.

The second concerned wider economic benefits which, as Chapter 8 indicates, is an area in which appraisal methodology is still being developed. There is emerging evidence on economic impacts (Quddus *et al*, 2004) but that still leaves open the question of the period of adjustment required and whether any economic changes could be attributed exogenously.

The third addressed changes in accessibility, which are reflected in changes in travel time and cost, but which may need to be presented in a disaggregate form as input to a distributional appraisal. While generalised cost may be an appropriate metric for accessibility, it raises the question of how road pricing charges themselves are perceived and valued.

The fourth covered the wider issues of equity and distributional impacts. The appraisal of distributional impacts generally is an area of weakness in appraisal methodology. To be informative to decision-makers the output needs to be kept disaggregate, and based on the estimated impacts on each of a number of predefined impact groups. Guidance is needed on how to select those impact groups, which may differ from one scheme to another, for example depending on the policy on exemptions. Care is also needed in deciding on how to present the disaggregate information on impacts, given that some (for example on low income groups) may be readily obtained from models, whereas others (e.g. on disabled travellers) will need to be more judgmental and qualitative.

The final issue was that of treatment of revenues generated, and their inclusion in benefit/cost ratios and estimates of value for money. Revenues may be treated as being a contribution to general public funds, in which case they can be treated at face value or assigned a shadow cost to reflect the value of their use in an optimal investment package. Alternatively they can be treated as hypothecated to a particular purpose (e.g.: the enhancement of public transport services), in which case it is preferable to appraise the package as a whole (Section 7.2.3). Guidance is needed on good practice in this area.

7.5.2 General limitations in appraisal of particular relevance to road pricing

Road pricing can be expected to have significant impacts on the reliability of journey times and, for public transport, on overcrowding. The valuation of improvements in reliability has

been the subject of considerable discussion (Bates *et al* 2001, Small *et al* 2005,). Levels of overcrowding will depend on the response of operators. This will depend on regulatory and infrastructure limitations and, with them, the application of the Mohring effect. Once again, the values to be assigned to time spent in overcrowded conditions are the subject of debate (Wardman, 2001).

Environmental benefits will include reductions in pollution and noise, for which resource values are increasingly available, and other more qualitative attributes such as enhancements in streetscape. These in turn might arise as a direct result of traffic reduction or through a policy of reallocation of the road capacity released. Methods such as hedonic pricing are available for valuing such benefits, but have not been widely applied.

One response to road pricing which is rarely predicted and is difficult to value concerns evasion of charges. This in turn requires an enforcement process whose costs and benefits need themselves to be appraised. Economic theory ignores transaction costs and there are no available methods at present for measuring the cost of compliance from the commuter's perspective. There is a clear link to scheme design here.

Finally, road pricing can be expected to have a significant impact on tax revenues, particularly through the reduction in fuel consumption. The principles adopted for deducting tax revenue losses from net benefits differ between countries, and can significantly affect the resulting benefits (Minken *et al*, 2003)