NATUR VÅRDS VERKET

SWEDISH ENVIRONMENTAL PROTECTION AGENCY

An instrument for assessing the quality of environmental valuation studies

An instrument for assessing the quality of environmental valuation studies

Orders
Phone: + 46 (0)8-505 933 40
Fax: + 46 (0)8-505 933 99
E-mail: natur@cm.se
Address: CM-Gruppen, Box 110 93, SE-161 11 Bromma, Sweden
Internet: www.naturvardsverket.se/bokhandeln

The Swedish Environmental Protection Agency
Phone: + 46 (0)8-698 10 00, Fax: + 46 (0)8-20 29 25
E-mail: natur@naturvardsverket.se
Address: Naturvårdsverket, SE-106 48 Stockholm, Sweden
Internet: www.naturvardsverket.se

ISBN 91-620-1252-5
ISSN 0282-7298

Print: CM Digitaltryck AB
Cover photos: Ablestock
Design: AB Typoform

Contents

Reading instructions

Do you want to assess the quality of a valuation study? Or do you need assistance in designing a valuation study? This report provides an instrument that will help you with these tasks. To fully understand how to use the instrument you need basic skills in environmental economics and economic valuation, including a basic knowledge of statistics/econometrics. Advanced skills should not be necessary. Note that the report does not replace economic valuation textbooks, rather it may usefully be complemented by modern valuation literature such as Bateman et al. (2002), Champ et al. (2003), Freeman (2003) and Haab and McConnell (2002). Some further references to relevant literature can be found in chapter 3, and the appendices of the report might also be of help to the reader. Use the report in the following way:

1. Read chapters 1, 2 and the introduction to chapter 3.

2. Read also appendices A, B1 and B2 if you need additional information on valuation methods and quality assessments of valuation studies.

3. If needed, read also additional literature. You will find suggestions for further reading in the report.

4. Download the electronic version of the evaluation form from www.naturvardsverket.se/bokhandeln/dse/620-1252-5 In this document, you fill in your answers to the instrument's check questions which relate to the quality of the study.

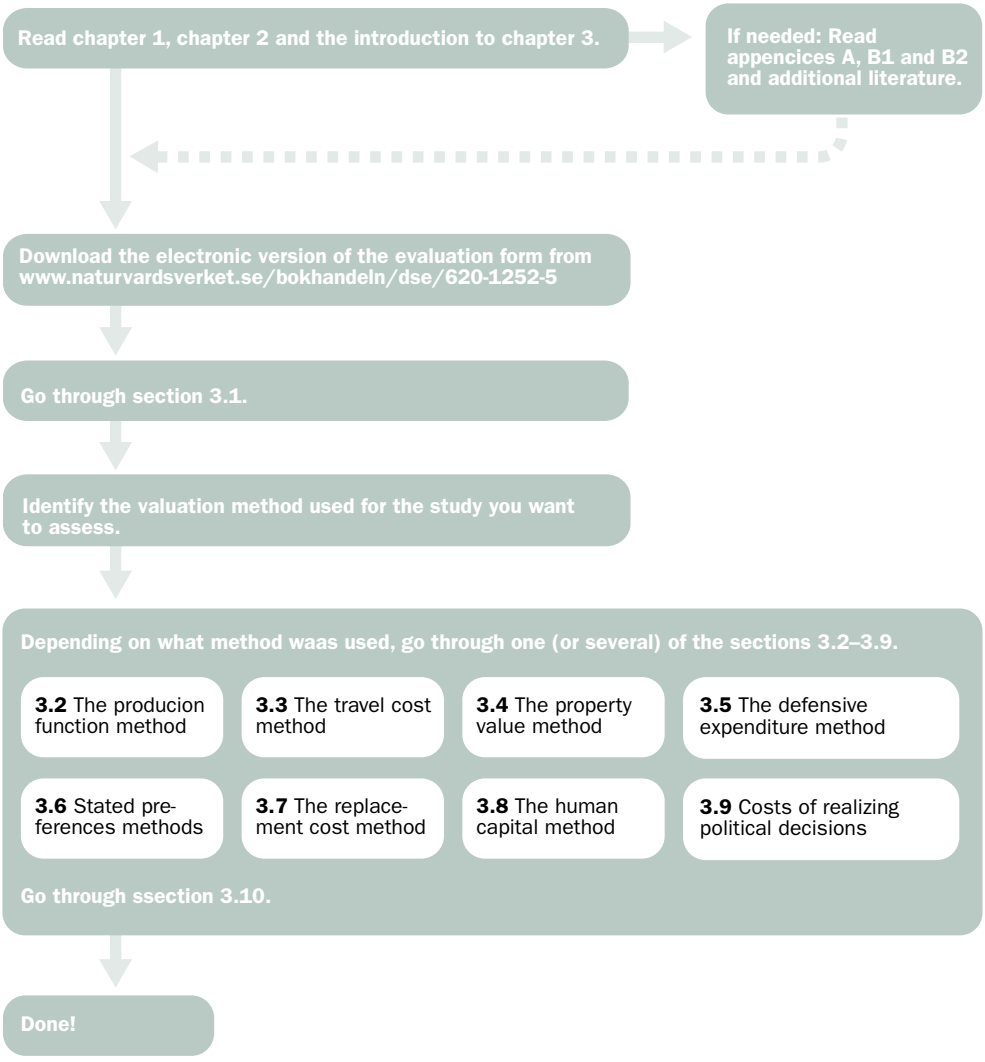
5. Go through section 3.1 and answer the check questions in that section.

6. Identify the valuation method(s) used for the study you want to assess.

7. Sections 3.2-3.9 contain check questions for each type of valuation method. Go through the relevant section(s) of the study you want to assess and answer the check questions.

8. Go through section 3.10 and make an overall assessment of the quality of the study.

9. Now you're done!



Foreword

In recent years, there has been an increasing demand for the inclusion of both benefits and costs in assessments of environmental policy proposals. However, difficulties in estimating the benefits side suggest that the positive effects of environmental policy risk being underestimated. One solution to this problem is to launch new valuation studies to increase the knowledge base in areas where few or no studies have been carried out to date. However, this requires a significant amount of time and financial resources. It is therefore important to use results from existing studies to the greatest possible extent. To this end, Valuebase^{SWE}, a Swedish database which includes more than 170 valuation studies, was set up in 2004 and there is also a handful of international examples of similar databases, e.g. EVRI. Whether results from existing studies should be used in analyses of new environmental policy proposals depends on the suitability and quality of the studies. The purpose of this report is therefore to provide an instrument that enables government agencies and consultancies to make consistent and clear assessments of the quality of existing valuation studies. The quality criteria in the report can also be of help in the design of new studies. We expect the instrument will help to improve the quality of economic analyses and thus provide a sound basis for environmental policy decisions.

The report was written by Tore Söderqvist and Åsa Soutukorva, Enveco Environmental Economics Consultancy. Their work was assisted and reviewed by a reference group consisting of researchers as well as representatives of government agencies: Fredrik Carlsson (Göteborg University), Per-Olov Johansson (Stockholm School of Economics), Bengt Kriström (Swedish University of Agricultural Sciences, Umeå), Daniel Thorburn (Stockholm University), Eva Samakovlis (National Institute of Economic Research), Sofia Grahn-Voorneveld (Swedish Institute for Transport and Communications Analysis), Anna Helena Lindahl (Swedish Environmental Protection Agency) and Håkan Marklund (Swedish Environmental Protection Agency). Oskar Larsson and Lars Drake managed the project on behalf of the Swedish Environmental Protection Agency.

The instrument was tested by desk officers from the target user group. The Swedish Environmental Protection Agency is grateful to these test pilots and to the members of the reference group for their valuable contribution.

Swedish Environmental Protection Agency, June 2006

1. Introduction

An increasing number of valuation studies

The number of empirical studies on the economic value of environmental change has increased rapidly during the last 20 years. For example, more than 5000 valuation studies from over 100 countries are included in a forthcoming bibliography (Carson, in preparation). The development is also evident in the establishment of databases of valuation studies and in the increasing number of introductory textbooks on economic valuation (e.g. Bateman et al. 2002, Champ et al. 2003). As regards Swedish studies, Kriström (1992) made a summary of approximately a dozen Swedish environmental valuation studies in the early 90's. Four years later Söderqvist (1996) summarised around 60 Swedish valuation studies, and recently 170 Swedish studies were compiled in a database called Valuebase^{SWE} (Sundberg and Söderqvist 2004a).

The increasing number of valuation studies reflect a general view that it is important and relevant to pay consideration to the environment and ecosystem services (cf. appendix A) in economic analyses, not least when designing and implementing policies. Such a need is expressed by the Swedish Environmental Protection Agency (Naturvårdsverket 2004) in a strategy proposal for the development of economic analysis in government agencies' environmental work. In the proposal it is emphasised that:

> "here are reasons to put further efforts into the development of methods for monetary measurement of environmental change, and to actually measure the value of environmental change in monetary terms." (p. 45).

An instrument for understanding and assessing the quality of valuation studies

If the results from valuation studies are to be used in a policy context, it is of great importance that the results are reliable. This is partly determined by whether or not the valuation studies are of an acceptable quality. The purpose of this report is to provide an instrument that is practicable in assessing the quality of valuation studies. The instrument is likely to increase the chances that valuation studies of good quality are used as a basis for policy decisions. The instrument identifies quality factors and thereby provides help to anyone who wants to evaluate a study; it points out which aspects the reader/user should pay attention to. However, quality is such a complicated feature that the instrument cannot be used for a simple grading of valuation studies. To convey an understanding for the complex nature of quality is another purpose of the instrument.

Whilst the main purpose of the instrument is to assist in assessments of existing valuation studies, it can also provide an understanding of what aspects are crucial to pay attention to when designing new studies. Hence, the instrument might be helpful for anyone who is planning to either carry out a valuation study or engage someone else to do valuation work.

The report is structured as follows:

- Chapter 2 provides a general discussion on which dimensions of quality might exist, and their relevance for valuation studies.
- Chapter 3 presents the instrument for assessing quality. The chapter identifies and discusses quality factors and contains questions associated with the factors. In order to facilitate filling in answers to the questions, there is a downloadable document template on www.naturvardsverket.se/bokhandeln/dse/620-1252-5

The reader will find additional information in the following appendices:

- Appendix A briefly describes the environmental economics methods that are available for valuing environmental change.
- Appendix B1 presents results from earlier work that has studied or discussed the quality of valuation studies. These concern earlier research, guidelines for carrying out valuation studies and how quality has been dealt with in valuation databases.
- Appendix B2 provides additional details about the conclusions of some selected studies on quality assessments of valuation studies.
- Appendix C is a glossary that includes some concepts that are defined in the report.

2. Quality dimensions of valuation studies

What is quality?

What is meant by quality? This basic question has to be answered before approaching the more specific task of assessing the quality of valuation studies. A very general definition of quality is ”fitness for use” (Juran and Gryna 1980). This definition suggests that the quality of something is dependent on what it is intended to be used for. Usefulness is also emphasised by SCB (2001a) in noting that the quality of a product is commonly viewed as being determined by the users' opinion of the product and its usefulness. This suggests that an assessment of the quality of a product should be based on product characteristics that are related to the extent to which the product fulfils needs and expectations among users (SCB 2001a). In what follows, four different dimensions related to the quality of valuation studies are discussed:

1. the user dimension – the preceding paragraph suggests that this dimension can be regarded as a kind of superior dimension,
2. the natural scientific-medical dimension,
3. the economic dimension, and
4. the statistical dimension

2.1 The user dimension

Can the study be used for what it is intended to be used for?

An important aspect of this dimension is that the quality of valuation studies is dependent on whether they actually can be used for what they are intended to be used for. Table 1 shows some important contexts in which valuation studies can be used. The comments in the table are made from a British perspective, but many of these contexts are found also in Sweden. For example, cost-benefit analyses including environmental aspects are carried out by some Swedish authorities, in particular the Swedish Road Administration, the Swedish Rail Administration and the Swedish Institute for Transport and Communications Analysis, and more Swedish authorities expect to carry out such cost-benefit analyses in the future (Frykblom and Helgesson 2002), see also SEPA (2004).

Is it possible for the user to make an objective quality assessment?

Another aspect related to the user dimension is the person who is supposed to assess the quality. When discussing and identifying quality criteria in this report, we assume that he/she has basic knowledge of economic valuation, but is not an expert in valuation. This point of departure implies that we to the greatest extent possible want to avoid that the person assessing the quality has to make subjective assessments. Our objective is instead to design quality criteria that are based on objectively observable study characteristics.

Table 1. *Some contexts in which valuation studies are used.*

Context	Comment from a UK perspective
Cost-benefit analysis: projects and programmes.	This is the context in which CBA was originally developed. Usually public investment projects in public or quasi-public goods.
Cost-benefit analysis: policies, including regulations.	In the UK, regulatory impact assessments are required for all regulations. Traditional for mainly regulatory impact assessments in the US.
'Demonstration' of the importance of an issue.	Usually used to estimate economic damage from some activity, e.g. behaviour towards health, pollution, noise.
Setting priorities within a sectoral plan.	Used for prioritising road investments.
Setting priorities across sectors.	Rare.
Establishing the basis for an environmental tax or charge.	Recent UK experience appears to be unique, e.g. landfill tax, possible pesticides tax.
'Green' national accounting.	Only utilised in minor way in the UK.
Corporate green accounting.	A few studies exist, but even fewer are public.
Legal damage assessment.	Not used in the UK but extensively used in the US.
Estimating discount rates.	Used in health literature and to derive discount rates in developing countries.

Source: Bateman et al. (2002).

2.2 The natural scientific-medical dimension

Is the valued environmental change realistic and relevant?

The valuation study has to rest on a sound natural scientific/medical basis related to the environmental change subject to valuation. The importance of such a sound basis is evident if the results of the valuation study are to be linked to an underlying environmental problem or policy. For example, if the purpose is to use the results of a valuation study in a cost-benefit analysis of measures against marine eutrophication, the valuation has to concern effects that can be accomplished by measures against the eutrophication.

Is it perceived in an objective way?

Another aspect related to the natural scientific-medical dimension is that an economic valuation is based on individuals' subjective perception of the envi-ronmental change subject to valuation. The willingness to pay is dependent on preferences and is thus subjective. But the subjective perception of an environmental change might be difficult to measure in an objective, scientific way. This is a

problem that often deserves attention. A typical example is how people perceive health risks. The defensive expenditure method (see Appendix A for a description) might give information about what individuals are willing to pay for measures reducing their health risks. However, the health risk reduction perceived by them might differ from the objective risk reduction. The way in which subjective risk reductions are translated to objective ones might be of critical importance for the result in a comparison of benefits and costs of risk reduction measures.

Another example of the implications of the difference between individual preferences and scientific knowledge might be difficulties for stated preferences (SP) methods (see appendix A for a description) to collect data solely about the values related to the environmental effects included in the valuation scenario. Individuals might have (more or less well-founded) opinions also about other effects that, according to them, would result if the scenario is realised, and it might be difficult to adjust for how these opinions influence the valuation.

2.3 The economic dimension

Does the study measure what it intends to measure?

A valuation study is not likely to have a high quality if it is unclear what the study aims at measuring. Economic theory gives a foundation for most of the valuation methods mentioned in appendix A, and these methods give – if they are properly designed – information on economic values in terms of the trade-offs that individuals/firms are willing to make for the sake of the environment. The methods thus estimate changes in wellbeing measured in ways that can be motivated by welfare economics, more exactly changes in the (Marshallian) consumer surplus, compensating variation or equivalent variation in the case of individuals, and changes in producer surplus in the case of firms. In contrast, the methods briefly described in section A.3 in appendix A are less consistent with economic theory.

It is often far from a matter of course to decide what measure of the change in individuals' wellbeing that should be estimated. The change in the Marshallian consumer surplus is from a theoretical point of view not fully satisfactory as a measure of wellbeing change. However, its weaknesses are not necessarily of importance in practice (Willig 1976), and it is evident that the change in the Marshallian consumer surplus is frequently used in practice in valuation studies when Marshallian demand functions are possible to estimate. Mainly in SP studies there are opportunities to design the study so that information is collected about compensating variation or equivalent variation. Whether information about compensating variation and equivalent variation are gathered by a question about willingness to pay (WTP) or willingness to accept compensation (WTA) depends on the direction of the environmental change, see table 2. The relevance of measuring compensating variation or equivalent variation is determined by, inter alia, how respondents perceive property rights (or moral rights) associated to the environmental change, see table 3.

Table 2. *The relationship between compensating variation and equivalent variation on one hand and questions about willingness to pay (WTP) and willingness to accept compensation (WTA) on the other hand.*

Measure of wellbeing change	Environmental improvement	Environmental deterioration
Compensating variation	WTP to obtain the improvement	WTA for the deterioration
Equivalent variation	WTA to forgo the improvement	WTP to avoid the deterioration

Source: See, e.g. Freeman (2003).

Table 3. *Compensating variation and equivalent variation interpreted in terms of property rights.*

Measure of wellbeing change	Environmental improvement	Environmental deterioration
Compensating variation	The individual has no right to the improvement (and thus has to pay to obtain it)	The individual has right to the initial situation (and thus has to be compensated for the deterioration)
Equivalent variation	The individual has a right to the improvement (and thus has to be compensated if it is not realised)	The individual has an obligation to accept the deterioration (and thus has to pay for preventing it)

Source: See, e.g. Freeman (2003).

Are the assumptions used in the study reasonable?

Some valuation methods are estimating economic values given strong assumptions, and these assumptions are not always reasonable. For example, the travel cost method and the property value method rely on the assumption that an environmental change only affects the wellbeing of the individuals actually using the environmental resource in question, i.e. the assumption of weak complementarity, see e.g. Freeman (2003). This can be illustrated by Swedish travel cost studies on environmental improvements in Stockholm Archipelago. These studies only estimate economic values associated to improvements for visitors to the archipelago. But people who (at least not at present) are not visiting the archipelago might very well also care about its environment. However, their willingness to pay for an improved archipelago environment cannot be captured by the travel cost study.

Conceptually, the total economic value of an environmental improvement might be divided into two components, use value and non-use value. A method relying on the assumption of weak complementarity is only estimating values associated to users. The values potentially held by non-users can only be captured by some SP method. It is thus reasonable for a valuation study to use an SP method if there are reasons to believe that there are substantial values held by non-users.

2.4 The statistical dimension

Were data collection, selection of statistical methods, aggregation to population levels etc., made in a reliable way?

The science of statistics usually concerns valuation studies in at least three ways:

1. When designing and carrying out the data collection, in particular if primary data are to be collected. Note that this work is also likely to take into account results from other disciplines, such as psychological findings about the effects of different ways of framing questions in a survey.
2. When selecting a method for statistical/econometric analysis of collected data and when carrying out the analysis. The choice of method might have a considerable impact on the results of the valuation study.
3. When aggregating value estimates to population levels. This procedure is strongly dependent on how the data collection was designed.

The recommendations for quality declaration of Swedish official statistics in SCB (2001a) illustrate what might be included in the statistical dimension. Besides information about the purpose of a statistical survey and who has commissioned it, the quality declaration should contain information about the contents, accuracy, timeliness, comparability, coherence, availability and clarity of the statistics. These requirements are summarised in table 4. An example of a quality declaration for Swedish official statistics is found in SCB (2001a). The recommendations are related to the functioning of a questionnaire in SCB (2001b).

Table 4. *Quality concept for Swedish official statistics.*

Main component A: Contents of the statistics.

This component concerns the statistical target characteristics. Subcomponents:

- Statistical target characteristics
  - Units and population
  - Variables
  - Statistical measures
  - Study domains
  - Reference times
- Comprehensiveness

Main component B: Accuracy of the statistics.

This component concerns the agreement between statistics and target characteristics. Subcomponents:

- Overall accuracy
- Sources of inaccuracy
  - Sampling
  - Frame coverage
  - Measurement
  - Non-response
  - Data processing
  - Model assumptions
- Presentation of accuracy measures

Main component C: Timeliness of the statistics.

This component concerns the relation of statistics to the current state of affairs. Subcomponents:

- Frequency
- Production time
- Punctuality

Main component D: Comparability and coherence of the statistics.

This component concerns how well different statistics can be used together. Subcomponents:

- Comparability over time
- Comparability between domains
- Coherence with other statistics

Main component E: Availability and clarity of the statistics.

This component concerns physical availability and intellectual clarity of statistics. Subcomponents:

- Dissemination forms
- Presentation
- Documentation
- Access to micro data
- Information services

Source: SCB (2001a).

The accuracy of statistics is a crucial quality component. It is determined by the extent to which different sources of error can be minimised. The sources of error for a statistical survey might be divided into errors caused by the fact that a sample is studied instead of a population (sampling error) and other errors (non-sampling error), which might arise because of the collection and processing of data. Sampling error is a deliberate consequence of statistical surveys because their basic idea is to use sampling for coming to conclusions about a population. Moreover, the consequences of sampling error are at least in principle possible to describe in detail by using confidence intervals for estimated parameters. A thorough theory is available which describes how this is done for different types of random samples, e.g. simple random sampling, stratified sampling, multistage sampling and cluster sampling, see, e.g. Cochran (1977). The situation becomes considerably less convenient when a non-random sampling procedure has been used, e.g. quota sampling or different types of convenience sampling where accessibility is determining the selection of respondents. Probability sampling is preferable when one wishes to know something about a population, and statistical surveys are supposed to make use of probability sampling procedures (Dalenius 1985, see also section 3). However, other type of samples might be justified in some situations. For example, being able to control who are selected to be included in the survey is sometimes more important than accomplishing a high degree of representativity of the population.

Non-sampling error is usually considerably less predictable than sampling error in a statistical survey. It might thus be difficult to find out the implications of non-sampling error, but this type of error is often likely to have a more negative effect on accuracy than sampling error (Biemer and Lyberg 2003). Table 5 presents five major sources of non-sampling error.

Model error is another important source of non-sampling error. This arises if the choice of statistical/econometric model is unsuitable for the intended estimation. For example, a serious model error might arise if a linear regression model is used for estimating the relation between two variables even if data indicate that the relationship is highly non-linear.

Table 5. *Five major sources of non-sampling error.*

1. *Specification error:* when the concept implied by the survey question and the concept that should be measured in the survey differ.
2. *Frame error:* when population elements are omitted or duplicated, or elements are erroneously included.
3. *Non-response error:* when there is unit non-response, item non-response or when responses to open-ended questions are incomplete.
4. *Measurement error:* when respondents deliberately or unintentionally provide incorrect information, interviewers fail to comply with the survey procedures, or questionnaires collect wrong information because of poor design.
5. *Processing error:* when errors occur in data editing, data entry or coding and when there are (human or software) mistakes in data analysis.

Source: Biemer and Lyberg (2003).

If a source of a non-sampling error is suspected to be present, it is important to try to find out if it causes a variable error or a systematic error, or both types of error (Biemer and Lyberg 2003). While variable errors increase the variance of estimates, the negative errors tend to cancel out the positive ones. This means that variable errors do not cause any bias in linear estimates such as estimated population means, population totals and population proportions. Variable errors and sampling errors thus affect linear estimates in a similar way. However, systematic errors result in biased linear estimates. As regards non-linear estimates, both variable and systematic errors might cause bias.

It exists a number of methods that can be used before or during the survey for reducing the presence of non-sampling errors. It is further possible to carry out analyses after the data collection with the purpose to find non-sampling errors and reduce their impact on the results. Table 6 presents some of these important methods and analyses.

Table 6. *Methods and techniques for reducing the presence of some types of non-sampling error.*

Stage of the survey process	Evaluation method	Purpose
Design	Expert review of questionnaire. Training of interviewers.	Identify problems with questionnaire layout, format, question wording, question order, and instructions. Increase chances of good interviewer performance.
Design/pre-testing	Cognitive methods, e.g. behaviour coding and cognitive interviewing.	Evaluate one or more stages of the response process.
Pre-testing/survey/post-survey	Debriefings such as interviewer group discussions or respondent focus groups.	Evaluate questionnaire and data collection procedures.
Pre-testing/survey	Observation, e.g. supervisor observation, telephone monitoring and tape recording.	Evaluate interviewer performance. Identify questionnaire problems.
Post-survey	Post-survey analysis, such as embedded experiments (e.g. variation in questions formats), non-random observation, tests of internal consistency and external validation. Post-survey data collection such as re-interview surveys and non-response follow-up studies.	Compare alternative methods of data collection. Estimate mean square error components, validate survey estimates.

After Biemer and Lyberg (2003).

2.5 Connections between the dimensions

The quality dimensions identified in the preceding sections constitute an attempt to sort out circumstances that are related to the quality of valuation studies. However, the dimensions are not independent of each other. This fact is illustrated by the examples of connections in table 7.

Table 7. *Some connections between the quality dimensions.*

	Use	Natural science	Economic theory
Natural science	Is there natural scientific knowledge detailed enough to allow comparisons between benefits and costs?		
Economic theory	How should estimated measures of changes in individual wellbeing be aggregated to population levels?	Are there big conflicts between natural scientific knowledge and individual preferences?	
Statistics	Accuracy of value estimates.	Accuracy of data on environmental change.	Collection of economic data. Estimation of measures of wellbeing change.

3. An instrument for quality assessment

About the instrument...

Factors of importance for quality

The purpose of this chapter is to provide a useful instrument for assessing the quality of valuation studies. The instrument involves an identification of a number of factors related to quality for...

a) ...valuation studies in general, irrespective of what valuation method was employed (section 3.1).
b) ...each of the valuation methods that are available (sections 3.2-3.9).

The quality of a valuation study is thus assessed partly through the quality factors in (a) and partly through the quality factors that according to (b) are relevant for the valuation method(s) used in the valuation study. Results reported in appendix B1, especially USEPA (2000), were used as a basis for identifying quality factors. Section 3.10 gives the user of the instrument an opportunity to give an overall assessment of the quality of the valuation study.

Check questions associated to each quality factor

Each quality factor is subject to a short description and discussion. Even if a quality factor can be identified, it is often difficult to operationalise the factor into a practical quality indicator. We make the operationalisation by using the description and discussion of quality factors as a basis for identifying one or several check questions. The great majority of these questions can be answered by an inspection of objectively observable characteristics of the valuation studies. The check questions are found in a table that in some cases is linked to a summarising motivation to why the questions are posed.

Most of the check questions can be answered by ”yes”, ”no” or ”don’t know”, and they were framed so that ”yes” answers are an indicator of good quality. Other check questions are instead about a piece of information associated with the quality of the valuation study, for example, the non-response rate. The question should in this case be answered by filling in text in the ”comment” column. From the viewpoint of quality, one situation when such pieces of information might be relevant is when comparing valuation studies for judging what study is most suitable for generalising valuation results to other settings (so-called benefit transfer). Note that some check questions are not relevant for some studies, and ”not relevant” should in such a case be written in the field for ”comment”. One example is that questions about the bid vector in a contingent valuation study are irrelevant if only open-ended WTP questions were used in the study.

For the sake of clarity, the check questions are numbered consecutively. Note though that this does not imply that all questions are to be answered when assessing a particular study. The questions in the sections 3.2-3.9 are associated with different valuation methods, which means that only the questions associated with the method(s) employed in the study are to be answered.

A document containing the check questions in sections 3.1-3.10 can be downloaded from www.naturvardsverket.se/bokhandeln/dse/620-1252-5. The reader may use this document as a form in which to fill in the answers to the questions.

Please note…

Supplementary comments

While the answers to the check questions should indicate the quality of the valuation study, it is also important that the user considers that assessing quality is not an easy task. Some of the difficulties should be clear from the description and discussion of the quality factors below. Moreover, a ”no” or ”don’t know” answer is not necessarily an indicator of bad quality. Whether it is so or not depends on the context. The last part of each of the sections 3.1-3.9 therefore consists of a field for filling in comments that supplement the answers to the check questions. For example, this field can be used for commenting on whether a ”no” implies a serious weakness of the valuation study or not.

The instrument gives you guidance, not a simple answer

To assess quality is a complicated task, and some of the questions are therefore likely to be difficult to answer. But what is really of importance here is not always to be able to give an unambiguous answer, but rather to obtain hints on what factors the user of the instrument should consider (or search for more information on) for getting an idea of the quality of the study. This means that the check questions are ”softer” than they sometimes might appear to be. Another reason for why it might be difficult to answer some questions is that valuation studies do not always include the pieces of information that are needed for finding an answer. This is a common problem when studies are published as journal articles. Strict space restrictions often imply that it is only possible to report the main result of the study. In such a case, a fair quality assessment might require that additional information about the study has to be collected. Journal articles often include references to one or several reports in which more detailed results can be found.

Usefulness is a relative term

The fact that assessing quality is complicated is also because quality is multidimensional. Further, the dimensions of quality are often intertwined. Four different quality dimensions were identified in chapter 2. Most of the check questions in the instrument are associated to the statistical, economic and natural scientific-medical dimensions. However, the questions are in some cases rather about the usefulness of the results of the valuation study. It is in this respect important to remember that usefulness is a relative quality because it depends on how the results are to be used. A limited usefulness is a problem only for those who therefore cannot make use of the results of the study.

3.1 Quality factors for all valuation studies

The following quality factors were identified as being relevant for all valuation studies irrespective of what valuation method the studies employed. The factors are explained in detail in below.

3.1.1 Earlier reviews
3.1.2 Principal/funder
3.1.3 Valuation method
3.1.4 Sensitivity analyses related to results from statistical/econometric analyses
3.1.5 Are future values discounted?
3.1.6 Primary data or secondary data?
3.1.7 Data collection
3.1.7.1 Survey, population and sample
3.1.7.2 The design of the data collection work
3.1.7.3 Data collection method
3.1.7.4 Non-response
3.1.7.5 Survey instrument
3.1.8 Access to data
3.1.9 Validity tests
3.1.10 Natural scientific/medical basis

3.1.1 EARLIER REVIEWS

The study might have been subject to one or several earlier reviews before it was finalised and reported. Such reviews are likely to have influenced its quality positively. Studies published in scientific journals have normally gone through a review of its scientific quality, which is an important indicator of good quality. However, such studies might not necessarily be useful in a policy context. Articles published in scientific journals are often about tests and development of methods. Value estimates from such studies might not be suitable to aggregate to a population level, maybe because a probability sample of respondents was not used for the study. On the other hand, there are studies that due to, for example, low scientific novelty, are not published in any scientific journal, but still are good applications of some valuation method. For a study which has not been published in any scientific journal, it is therefore important to find out if it still has been subject to some kind of external review. Non-published parts of PhD theses are an important example of such studies. Other examples might be licentiate theses, master theses and agency reports whose production has involved an external reference group.

Earlier reviews of the study *should affect the quality of the study positively. However, the review might have been more or less thorough.*

Check questions	Yes/no/don't know	Comment
1. Has the study been subject to external review?		
1a. If "yes", in what way?		

3.1.2 PRINCIPAL/FUNDER

The results of a valuation study might be used for promoting the realisation (or the prevention) of projects. It can therefore not be precluded that valuation studies are designed in a biased way. This implies that is important to know who was conducting the study and who was the principal/funder.

Is there any risk *of biases because of those who conducted and/or funded the study?*

Check questions	Yes/no/don't know	Comment
2. Who conducted the study?		
3. Who commissioned/funded the study?		

3.1.3 VALUATION METHOD

There are a number of valuation methods available for economic valuation of environmental change, see appendix A. Some of them are designed for measuring changes in consumer surplus and/or producer surplus and can thus be motivated from the viewpoint of welfare theory. Such methods include:

- The production function method (PF)
- The travel cost method (TCM)
- The property value method/hedonic price method (HP)
- The defensive expenditure method (DE)
- The contingent valuation method (CVM)
- Choice experiments (CE)

Specific quality factors for these methods are identified in sections 3.2-3.6.

Other valuation methods are not equally well founded in welfare theory. While this does not preclude that they produce useful information, it is a weakness because a more vague theoretical basis might make it difficult to interpret the valuation results. The following methods are found in this group:

- The replacement cost method (RCM)
- The human capital method (HCM)
- Costs of realising political decisions ("political WTP", pWTP)

Specific quality factors for these methods are identified in sections 3.7-3.9.

A valuation study typically makes use of one of these valuation methods. However, sometimes two or more methods are used in the same study. For example, it happens that the travel cost method is combined with the contingent valuation method.

Valuation methods *based on welfare economics have a clear theoretical basis. This facilitates the interpretation of results from applications of these methods.*

Check questions	Yes/no/don't know	Comment
4. What valuation method was used?		
5. Is the valuation method rooted in welfare economics?		

3.1.4 SENSITIVITY ANALYSES RELATED TO RESULTS FROM STATISTICAL/ECONOMETRIC ANALYSES

One of the main difficulties associated with interpreting results from valuation studies is due to the fact that the choice of statistical/econometric method for analysing data might have a substantial impact on the size and uncertainty of the estimates. A good study is expected to report the statistical uncertainty in terms of, for example, confidence interval or standard deviations, but the dependence of statistical uncertainty on the choice of statistical/econometric method implies that information on statistical uncertainty is not sufficient for assessing the total uncertainty. Moreover, considerable knowledge of economics and statistics/econometrics is generally required for judging whether the choice of method for analysis was reasonable, given such things as the structure of the data collected. The difficulty to judge whether the choice of method was reasonable and to know the impact of the choice of method on the size and uncertainty of estimates suggests that valuation studies should include different types of sensitivity analyses. Sensitivity analyses indicating what could reasonably be a lower and upper boundary for the valuation estimates would be particularly helpful. This is because information on the lower and upper boundaries can be sufficient for making conclusions in a cost-benefit analysis if the costs of the project in question are smaller than the lower boundary or greater than the upper boundary. For example, such a sensitivity analysis might show the consequences of using alternative (but reasonable) methods for statistical/econometric analysis and using alternative (but reasonable) assumptions in a given method, for example, concerning the choice of probability distribution. Considerable knowledge of economics and

statistics/econometrics is again needed for judging what alternatives are reasonable, and for a basic quality assessment it has to be taken for granted that the authors of the study have made a good judgment of what is reasonable and not reasonable.

Estimates of economic *values often have uncertainties attached to them. A basic way to report uncertainty is to use statistical measures such as confidence intervals and standard deviations, and it is important to know how big this uncertainty is. For example, is the estimated value significantly different from zero? However, there are other types of uncertainties that such statistical measures do account for. It is therefore desirable to also have a broader sensitivity analysis which indicates the lower and upper boundaries of the economic values.*

Check questions	Yes/no/don't know	Comment
6. Was the statistical uncertainty of the estimated economic values reported in terms of, for example, confidence intervals or standard deviations?		
6a. If "yes", fill in the estimated economic values and their associated uncertainty.		
7. Was there a sensitivity analysis indicating what is reasonably the lower boundary of the estimated economic values?		
7a. If "yes", fill in this lower boundary.		
7b. If "yes", what factors were considered in the sensitivity analysis?		
8. Was there a sensitivity analysis indicating what is reasonably the upper boundary of the estimated economic values?		
8a. If "yes", fill in this upper boundary.		
8b. If "yes", which factors were considered in the sensitivity analysis?		

3.1.5 ARE FUTURE VALUES DISCOUNTED?

It is not unusual that a valuation study estimates economic values that are realised in the future. One example might be benefits to farmers because of a water quality improvement. The effects of the improvement might take time, so that farmers' producer surplus is not affected until a number of years has passed. When time enters in the analysis, there is a need to convert future values into present values. This is usually carried out by a discounting procedure in which the choice of discount rate can have a great impact on the size of present value. It is therefore important that the valuation study reports on how the present value calculation was carried out and how the choice of discount rate was motivated. In the scientific debate about discounting, it is possible to discern two different approaches to discounting: a descriptive approach arguing that the actual behaviour at capital markets should determine the size of the discount rate, and a prescriptive approach arguing that ethical considerations should be the basis for selecting a discount rate; see, e.g. Arrow et al. (1996). The presence of different approaches indicates that the choice of discount rate should not be made in a routine manner.

Check questions	Yes/no/don't know	Comment
9. If the valuation study estimated future economic values, did the study report how these values were converted into present values?		
9a. If "yes", how was the selected discount rate motivated?		
9b. If "yes", what was the size of the discount rate that was used?		

3.1.6 PRIMARY DATA OR SECONDARY DATA?

Data of good quality play a decisive role for the reliability of the results of a valuation study. Data can either be primary or secondary data. The former refers to data that were collected with the purpose of being used for the valuation study in question, and the latter is data that were collected earlier in some other context.

How to handle secondary data?

The quality factors in section 3.1.7 below are about the collection and preparation of primary data. A study using secondary data does probably not include enough information on the original data collection for making it possible to answer the check questions in section 3.1.7. But a study using secondary data should still contain an evaluation of how data once were collected. Such an evaluation should consider the issues that are brought up in section 3.1.7. Even if the check questions cannot be answered for a secondary data study, the text and

questions in section 3.1.7 can thus still be helpful for judging the evaluation of the original data collection that a secondary data study should include.

A potential weakness associated with secondary data is that the main purpose of the original data collection might not have been to collect the particular data that were used in the valuation study. If this is the case, there is a risk that the original data collection involved relatively small efforts for ensuring a high quality of these particular data. Moreover, to decide to what degree secondary data are suitable for being used in a new study is often a matter of judgment. For example, the original data collection might have concerned another population (e.g. the US population), but the data collected was still judged to be sufficiently relevant for the population of interest to the valuation study (e.g, the Swedish population), possibly after adjustments for known differences among the populations.

Primary data *are likely to be more suitable for the purpose of the valuation study. The original data collection should have been evaluated if secondary data were used.*

Check questions	Yes/no/don't know	Comment
10. Were primary data used?		
11. If secondary data were used, was the quality of the original data collection evaluated?		
11a. If "yes", what was the result of this evaluation?		
12. If secondary data were used, was the main objective of the original data collection to collect the data that were used in the valuation study?		
13. If secondary data were used, was the relevance of using it for the valuation study evaluated?		

3.1.7 DATA COLLECTION

This section is primarily intended for studies using primary data, but it might also be helpful for assessing an evaluation of data quality in a study using secondary data, cf. section 3.1.6.

3.1.7.1 Survey, population and sample

It is generally an advantage if the data collection was carried out as a survey. Table 8 contains general prerequisites that together define a survey. It might

Table 8. *Criteria that together define a survey.*

Criterion	Comments
1. A survey concerns a set of objects comprising a population.	Defining the target population (i.e. the population of interest) is critical both for inferential purposes and to establish the sampling frame.
2. The population under study has one or more measurable properties.	Those properties that best achieve the specific goal of the project should be selected.
3. The goal of the project is to describe the population by one or more parameters defined in terms of the measurable properties.	Given a set of properties, different parameters are possible, such as averages, percentiles, and totals, often broken down for population subgroups.
4. To get observational access to the population, a frame is needed, i.e. an operational representation of the population units, such as a list of all objects in the population under study or a map of a geographical area.	It is often difficult to develop a frame that covers the target population completely.
5. A sample of objects is selected from the frame in accordance with a sampling design that specifies a probability mechanism and a sample size (i.e. a probability sample).	The sampling design always depends on the actual circumstances associated with the survey. For example, skewed populations may require stratified sampling. Every sampling design must specify selection probabilities and a sample size.
6. Observations are made on the sample in accordance with a measurement process.	Data collection can be administered in many different ways. Often, more than one mode must be used.
7. Based on the measurements, an estimation process is applied to compute estimates of the parameters when making inference from the sample to the population.	The error caused by a sample being observed instead of the entire population can be calculated by means of variance estimators. The resulting estimates can be used to calculate confidence intervals. However, not all the errors in the survey data are reflected in the variances.

Source: Dalenius (1985), Biemer and Lyberg (2003).

be difficult or impossible to use sample data for coming to conclusions about aggregate economic values for a population if any of these prerequisites is not satisfied. For example, probability samples are sometimes not used, which means that selection probabilities are not known for all objects (e.g. individuals or households) in the population. Self-selection is another common problem that implies that a data collection cannot be classified as a survey. An example of self-selection might be a travel cost study collecting data on visits to a recreational area by placing questionnaires in cabins in the area. Besides the problem that the questionnaire is only found by those visiting a cabin, it is probably only visitors who are interested in the questions that fill in the questionnaire. Probability sampling should be chosen whenever representativity for a population is a desirable

feature, which is often the case. However, non-probability sampling might be adequate in some situations, for example, when representativity is judged to be less important than being able to control who are included as objects in the study.

The minimum sample size necessary for obtaining a desired degree of certainty in population estimates depends on the degree of variability associated with the variables of interest to the valuation study. It is therefore not possible to identify a generally valid minimum sample size for valuation studies. However, one might note that samples used in Gallup polls with the aim of saying something general about the attitudes among Swedish adults usually consist of at least 1000 individuals. Carson (2000) recommends a sample size of at least 300-2000 objects for CVM studies.

The check questions below focus on three crucial survey features: the definitions of a target population and a sampling frame, and the sampling method. See Svenska Statistikersamfundet (2005) for recommendations on how populations and samples should be described.

The target population is the population that the study actually wants to come to conclusions about, whereas the frame population is the population that in fact was used as a basis for the survey. One option is to study all objects in the frame population, but since this in most cases is a too expensive option, it is more common to draw a sample instead. A number of objects is then selected from the frame population, which in this case constitutes the so-called sampling frame.

The frame population/sampling frame often differs from the target population. This might be due to practical reasons. There might not be directories or registers available that perfectly cover the objects in the target population. This can result in overcoverage, i.e. there are objects that are included in the frame population, but not in the target population, and/or undercoverage, i.e. there are objects in the target population that are not included in the frame population.

For example, the target population might have been defined as all individuals living in a city, but a study might choose to limit the target population to all individuals domiciled in the city because it is possible to get access to a census register. In this case, all individuals who live in the city without being domiciled there are excluded from the study (undercoverage), whereas all individuals who are domiciled in the city but in fact lives somewhere else are included (overcoverage). It might be important to take such potential differences between the target population and the frame population/sampling frame into account.

Since it is not possible to identify a generally valid minimum sample size for valuation studies, only one check question is posed about the sample size. If the valuation study estimated aggregate economic values for the population, it is important that the way of computing these estimates is consistent with the definition of the population and the sampling procedure. For example, if the probability of being selected to the sample varied among different population groups (e.g. in the case of stratified sampling), this has to be taken into account in the computation of estimates for the population.

A valuation study *aiming to estimate values that are representative for a population should be designed as a survey. Crucial issues in such a design include the definitions of target population and sampling frame, and the use of probability sampling for constructing a sample. A survey might not be necessary if the valuation study has some other purpose, e.g. carrying out some test of a valuation method. Check questions 14-20 are about some important aspects of a survey. Question 21 provides a possibility to make an overall judgment on the basis of table 8.*

Check questions	Yes/no/don't know	Comment
14. Was a target population defined?		
14a. If "yes", how was the target population defined in time and space, and what was its size?		
15. Was a frame population/ sampling frame defined?		
15a. If "yes", how was the frame population/sampling frame defined in time and space, and what was its size?		
16. Were potential differences between the target population and the frame population/ sampling frame reported?		
17. How did the study take into account potential differences between the target population and the frame population/ sampling frame?		
18. What was the sample size?		
19. What type of sampling procedure was used for constructing the sample?		
20. Was the sampling procedure a probability sampling?		
21. On the whole, did the study meet the criteria that define a survey?		

Check questions	Yes/no/don't know	Comment
22. If "no" to question 21, was the purpose of the study of a kind that does not motivate a survey? (For example, it might not be necessary to carry out a survey if the study was not aiming at computing estimates which are representative for a population.)		
23. If aggregate economic values for a population were estimated, was this estimation consistent with the sampling procedure and the definition of the population?		

3.1.7.2 The design of the data collection work

The data quality is also determined by the design of the data collection work. There are several methods and principles available for questionnaire design and the implementation of interview and mail questionnaire studies. For example, CVM studies have often employed methods developed by Don Dillman, such as the total design method and the tailored design method (Dillman 1978, 1991, 2000). However, these methods were developed in a US context, and a Swedish valuation study should also consider Swedish experience (e.g. SCB 2001b, Wärneryd 1990). One way of ensuring that sound methods are used is to involve an expert in data collection in the study. Further, survey instruments such as mail questionnaires should be developed and tested by using focus groups (or the like) and a pilot study. Weaknesses in the design of the data collection work might result in, for example, a substantial non-response rate, cf. section 3.1.7.4.

Check questions	Yes/no/don't know	Comment
24. Did the valuation study involve any experts in data collection?		
25. Were focus groups (or the like) consulted when developing and testing the survey instrument?		
26. Was a pilot study carried out to test the survey instrument?		

3.1.7.3 Data collection method

Face-to-face interviews, telephone interviews and mail questionnaires are traditional methods for data collection, but computer technology including the Internet and e-mail has introduced other methods. Some main methods available are found in table 9. The table divides the methods according to the degree of contact with respondents and the degree of data collector involvement. Note that methods might be combined. For example, a telephone interview can be preceded by mailing questions and information to the object. A data collection option not mentioned in the table is to distribute a questionnaire to a group of individuals who are asked to fill it in on the spot. A data collector is however available all the time for answering questions that the respondents might have.

Table 9. *Data collection methods.*

Degree of contact with respondent	High data collector involvement		Low data collector involvement	
	Paper	Computer	Paper	Computer
Direct	Face-to-face (paper-and-pencil interviewing)	Computer-assisted personal interviewing)	Diary	Computer-assisted self-interviewing
Indirect	Telephone (paper-and pencil interviewing)	Computer-assisted telephone interviewing	Mail, fax, e-mail	Touch-tone data entry, e-mail survey, web, disk by mail, voice recognition entry
None	Direct observation	Computer-assisted data entry	Administrative records	Electronic data interchange

Source: Biemer and Lyberg (2003).

What method should be used? The answer depends on many different factors. According to Arrow et al. (1993), CVM studies should not make use of mail questionnaires. Carson (2000) emphasises that face-to-face interviews increase the chance that respondents understand the scenario, since such interviews facilitate the use of visual aids such as photographs, drawings, maps, etc. In our opinion, it is not possible to come to a general conclusion about what data collection method is the best one, but the choice of method is dependent on the context. For example, face-to-face interviews are expensive, and the presence of an interviewer might result in biases due to phenomena such as a tendency that respondents give answers that they believe please the interviewer. On the other hand, face-to-face interviews are characterised by a great flexibility and tend to result in a high response rate. Telephone interviews are less expensive than face-to-face interviews, but telephone interviews have to be shorter and there is

no possibility to use visual aids unless such material is sent to the respondent in advance. However, using the Internet for questionnaires provides an opportunity to include visual aids. Biases due to the presence of an interviewer are avoided by using mail questionnaires. Further, mail questionnaires are probably more suitable for collecting information about sensitive issues and are relatively cheap to use, but they might result in a low response rate, not least low item response rates. Some factors that are of importance when selecting data collection method are listed in table 10.

Table 10. *Some important factors to consider when selecting data collection method.*

Factor	Implication for mode choice
Concepts to be measured	If a visual medium is required, a telephone survey can be ruled out. Complex concepts usually benefit from interviewer assistance.
Target population to be surveyed	Can the non-telephone population be ignored? If so, consider the telephone mode. Literacy level: Mail modes require literacy rates at or above the national average. What language(s) should be used? Does the target population include a large proportion of immigrants, foreign visitors, etc.?
Contact information available on frame	If name and address are available, mail or face-to-face interview should be considered.
Saliency of the topic	If much persuasion is needed to obtain adequate response rates, mail surveys must be ruled out.
Speed of completion	If needed very quickly, telephone is best. If needed in weeks, a mail survey may be feasible.
Scope and size of the sample	For a national survey, cost may be the reigning factor that suggests mail or telephone survey.
Sample dispersion	Maximum dispersion suggests a mail or telephone survey. In face-to-face surveys, some clustering is almost always needed.
Frame coverage of target population	If only poor coverage frames are available, use a face-to-face survey, random digit-dialing, or mixed-mode.
Non-response	Interview modes usually generate higher response rates than self-administered. Ability to persuade reluctant sample units depends on richness of media (e.g. in mail surveys, motivation is limited to written materials). Non-response is confounded with coverage problems in mail and telephone modes. Mail questionnaires might be regarded as junk mail and thrown away by sample units.

Factor	Implication for mode choice
Interviewer	Interviewer can generate response errors, such as social desirability bias. Interviewer-assisted mode is not good for collecting sensitive information. Interviewer necessary for visual aids and probing. Centralised telephone interviewing reduces costs and errors compared to non-centralised interviewing. Telephone interviewers can have larger work-loads due to no travel burden.
Respondent	There is some evidence that respondents prefer self-administered surveys. Self-administered modes are suitable for collecting sensitive information. If the response task is difficult, interviewer assistance is necessary.
Instrument	Mail questionnaires must be relatively simple but are suitable when questions contain many response alternatives. Complex instruments call for the interview mode. Mixed-mode must use questionnaires that can be used in all modes.
Cost	Everything else may be secondary if mail is the only mode that can be afforded.

Adapted from Biemer and Lyberg (2003).

A check question is posed below about when the data collection was carried out. Valuation methods are refined over time and people's preferences (and income) change, which suggest that relatively new studies have an advantage over older studies, other things being equal. Information about the point of time might also be helpful for judging whether the data collection was carried out when media paid considerable (or little) attention to the environmental quality subject to valuation.

Whether a suitable *data collection method was used or not has to be judged from case to case. Important factors that determine what method is suitable include the following: Are the questions very complex? Is it necessary to communicate a lot of information to the respondents? It is important to know when the data collection was carried out to be able to judge whether the data are out-of-date or not.*

Check questions	Yes/no/don't know	Comment
27. What data collection method was used?		
28. When was the data collection carried out?		

3.1.7.4 Non-response

Non-response might reduce the reliability of the collected data. Non-response refers to the phenomenon that values are missing for one or several variables that a study is aiming at collecting information on. There are two main types of non-response:

1. Unit non-response: All values are missing for the object in question (e.g. when an individual has not at all answered a mail questionnaire).
2. Item non-response: Only some values are missing for the object in question (e.g. when an individual has not answered some of the questions in a mail questionnaire).

It is not possible to identify any general rule for what unit non-response is the maximum acceptable one. According to Carson (2000), a non-response of 20-40% is small. However, a 25% non-response in one study might be more serious than a 40% non-response in another one. Whether non-response has serious consequences or not does not only depend on the non-response rate, but also on how respondents and non-respondents differ with respect to the variables of interest to the study, for example, willingness to pay. We therefore do not recommend any rule stating that, for example, 50% is the maximum acceptable non-response rate, but that the non-response rate is reported as it is, and is supplemented with information about how the study has handled the non-response. Note that non-response and response rates can be defined in several different ways (Biemer and Lyberg 2003, p. 86). Svenska Statistikersamfundet (2005) recommends what measures of response and non-response should be used, and also suggests that a report on response and non-response should include:

- Number of respondents giving usable observations.
- Number of sampled objects not giving usable observations.
- Choice of one or several non-response measures and their numerical values.
- The size of the systematic errors that the non-response might have caused.
- Measures that were taken for reducing the effects of systematic errors.

Further, Japec et al. (1997) recommend that the non-response report should include the following items:

- The size of unit non-response for different types of objects related to the sample and the population.
- The extent of item non-response for important variables.
- Reasons for non-response.
- Measures that were taken for reducing non-response, e.g. a follow-up study of non-respondents.
- An assessment of how non-response affects the results of the study.
- Methods for adjusting for non-response in estimation procedures.

Valuation studies often assess the effects of non-response on the results of the study by making more or less extreme assumptions about the willingness to pay of non-respondents, for example, that non-respondents have a zero WTP. However, it should be noted that more advanced methods for handling and analysing non-response are available; imputation and weighting are two principal methods. See Lundström and Särndal (2001) for an introduction.

Non-response might cause *unreliable results. Potential systematic differences between respondents and non-respondents should be taken into account when estimating aggregate economic values for a population.*

Check questions	Yes/no/don't know	Comment
29. Was there a report on non-response?		
30. How was unit non-response defined?		
31. What was the size of the unit non-response (in percent)?		
32. Was a follow-up study of non-respondents carried out?		
33. According to the study, how are valuation results affected by the non-response?		
34. If values at a population level were estimated, did such estimations take non-response into account?		

3.1.7.5 Survey instrument

The valuation study should include a copy of the survey instrument that was used. For example, if a mail questionnaire was carried out, the study should contain a copy of the whole questionnaire, including all information that was communicated to the respondents, e.g. cover letter, valuation scenario and facts about the environmental change subject to valuation. However, space limitations might imply that copies of the complete survey instrument cannot be included in some publications. In such a case, the survey instrument should instead be available in a background report or the like.

It is sometimes difficult to report the whole survey instrument because of the choice of data collection method. If applicable, this should be mentioned in the "comment" field for the check question below. For example, computerised questionnaires might include features that are difficult to reproduce in a publication.

Check question	Yes/no/don't know	Comment
35. Was a copy of the complete survey instrument presented along with the results or otherwise available?		

3.1.8 ACCESS TO DATA

It is an advantage if it is possible to get access to the data used. This gives an opportunity to make other analyses than those carried out in the valuation study. It might also make it possible to carry out meta studies, where data from several different studies are merged in order to make more general analyses.

Check question	Yes/no/don't know	Comment
36. Did the study mention whether it is possible to get access to the data used?		

3.1.9 VALIDITY TESTS

One way of checking data quality is to design the data collection in a way that enables tests of internal validity. This can be done in many different ways dependent on what data are collected. If primary data are collected, one common way to test for internal validity is to include two or several questions whose answers should confirm each other. Such a test should at least concern data that are crucial for the valuation study, i.e. the data that are used for estimating economic values. The nature of these data varies from case to case. For example, an important part of a travel cost study might be collection of data on the consumption of the environmental quality that is to be valued, for example, data on fish catches if the study is about the recreational value of improved recreational fishing. In such a study, it might be wise to complement the main questions about fish catches with a question whose answers should be consistent with the results of the main questions. Internal validity can also be tested in studies using secondary data. For example, a register might include different types of income data that should confirm each other. Some valuation methods involve specific tests of internal validity, and check questions about such tests are found in later sections. One important example is scope tests in stated preferences studies; see section 3.6.1.

It is also desirable to test whether the results of the valuation study show external validity. One basic way to do this is to compare the value estimates to estimates from earlier, similar studies. However, it might also be relevant to make such comparisons for other types of estimates. Sticking to a recreational fishing

example, it could be a good idea to compare the fish catch estimates of the travel cost study to such estimates from earlier studies (if any).

Check questions	Yes/no/don't know	Comment
37. Was there any test of internal validity?		
37a. If "yes", what test was carried out?		
37b. If "yes", did the test indicate the presence of internal validity?		
38. Was there any test of external validity?		
38a. If "yes", what test was carried out?		
38b. If "yes", did the test indicate the presence of external validity?		

3.1.10 NATURAL SCIENTIFIC/MEDICAL BASIS

The valuation study is likely to have benefited from advice from experts in natural sciences/medicine. An involvement of such experts increases the chances of a sound natural scientific/medical basis for the valuation study.

Check question	Yes/no/don't know	Comment
39. Were any experts in natural sciences/medicine involved in the valuation study?		

3.1.11 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.2 Quality factors for the production function method

The following quality factors were identified for the production function method:

3.2.1 Natural scientific basis
3.2.2 Estimation of changes in producer surplus
3.2.3 Modelling of the whole market including dynamic effects

3.2.1 NATURAL SCIENTIFIC BASIS

The point of departure for an application of the production function method is presence of knowledge of how an environmental change affects the production of a good or service. The objective is to estimate a production function in which the environment/ecosystem service is included as one of several inputs. Detailed natural scientific knowledge is often required for accomplishing such an estimation. For example, agronomical knowledge is needed if the valuation study is about ecosystem services influencing farmers' crop production. It is therefore likely to be important for a study using the production function method that a "yes" answer can be given to check question #39, i.e. that experts in natural sciences/medicine have been involved in the study. Further, it is probably an advantage if the natural scientific knowledge of the relation between an ecosystem service and the production of a good is based on an explicit cause-and-effect modelling and not only on statistical correlations from, for example, time series data on the supply of the ecosystem service and the production of the good.

Check questions	Yes/no/don't know	Comment
40. Is the relation between the ecosystem service and the production of the good in question studied in a cause-and-effect way?		
40a. If "yes", how was this relation studied?		

3.2.2 ESTIMATION OF CHANGES IN PRODUCER SURPLUS

It is important that the study does not only estimate changes in producers' revenues or costs because of an environmental change, since this only in special cases gives information on changes in producer surplus. In general, estimation of changes in producer surplus requires knowledge of how producers react to the environmental change by, for example, adjustments of the use of other inputs.

Check question	Yes/no/don't know	Comment
41. Were changes in producer surplus estimated?		

3.2.3 MODELLING OF THE WHOLE MARKET INCLUDING DYNAMIC EFFECTS

It is an advantage if the study includes a modelling of both supply and demand for the good whose production is affected by the environmental change. This makes it possible to estimate changes in both producer surplus and consumer surplus. It might also be important to apply a general equilibrium analysis or other tools for analysing the presence of potential dynamic effects.

Check questions	Yes/no/don't know	Comment
42. Was the whole market for the produced good modelled?		
43. Was there any analysis of potential dynamic effects at the market for the produced good and related markets?		

3.2.4 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not

3.3 Quality factors for the travel cost method

The following factors were identified as particularly important for assessing the quality of travel cost studies:

3.3.1 Definition of site(s)
3.3.2 Sampling strategies
3.3.3 Model specification
3.3.4 Calculation of travel costs
3.3.5 Opportunity cost of time
3.3.6 Multipurpose trips
3.3.7 Selection of environmental quality variable

The original (zonal) travel cost method is nowadays used rather seldom. Choices of recreational sites are instead normally modelled by using a random utility model (RUM) and individual data. If not otherwise stated, the quality factors and check questions below are valid for both the zonal method and the RUM based method.

3.3.1 DEFINITION OF SITE(S)

The recreational site is defined as a park, beach, lake, etc. in the zonal travel cost method. The site is often an area with clearly defined geographical borders, but in cases when the delimitation is less clear, the site has to be defined by the analyst. A good study quality is likely to require a reasonable definition. Whether the definition of the site is reasonable or not depends on, for example, if it is possible to include the whole area that is affected by a particular policy. For instance, assume that the destination of the travel is a marine reserve in a coastal area. It would probably be unreasonable in such a case to define the site as only a part of the reserve.

This check question *is applicable to studies which used the zonal travel cost method.*

Check question	Yes/no/don't know	Comment
44. How was the site defined?		

In a RUM model, a number of sites are defined and included in a choice set, i.e. the sites that the respondents choose among. The sites in such a set might be, for example, all lakes larger than 100 hectares within an area (Parsons and Kealy 1992), or the most important climbing areas within a region (Shaw and Jakus 1996). Sites can also be different types of administrative areas (Andrews 1996). The definition of a site might thus vary from very large areas to very narrowly defined places, and the number of sites might vary from just a few to several

thousands. In practice, the selection of sites in the choice set is mostly made by the analyst. However, there is a risk that the sites chosen by the analyst do not reflect the real choice set of the respondents, which suggests that it might be an advantage to let the respondents state what sites they chose among. Also the quality of RUM studies is dependent on how sites were defined. For example, it is likely to be inappropriate to not include places suitable for bathing in the choice set if the study aims at valuing improved bathing water quality in a coastal area.

These check questions *are applicable to studies which used the RUM based travel cost method.*

Check questions	Yes/no/don't know	Comment
45. How was the choice set defined?		
46. Did the study include any effort to find out respondents' actual choices among sites?		

3.3.2 SAMPLING STRATEGIES

An important phase in a zonal travel cost study is the sampling of individuals whose travel behaviour is to be studied. Sometimes the sample is limited to the population of visitors to the site, but some studies extend the population to non-visitors. While it is generally an advantage to take non-visitors into account in the analysis, there might be situations when only visitors are of interest. If only visitors are included in the analysis, it is important that the estimated recreational values are not interpreted as being valid also for non-visitors. In the case of RUM based studies, sampling is usually not restricted to visitors, but the sample is almost always made in other ways, such as random sampling of the general public. For both types of RUM studies, the sampling procedure is an important factor to take into account because it might affect the estimates of recreational values.

Check question	Yes/no/don't know	Comment
47. How was the sampling designed with respect to visitors and non-visitors?		

3.3.3 MODEL SPECIFICATION

In a zonal travel cost study, the demand for recreation to a specific site is analysed. Besides travel costs, income and variables related to quality, it is possible to take the presence of potential substitute sites into account when specifying the model by, for example, including travel costs to these sites. In a RUM model, site-specific variables are crucial because the respondents' choices of sites depend on, inter alia, the environmental quality at the different sites. The quality of a RUM study is probably dependent on the availability of site-specific data such as services, lodging options and communications. Check question #48 is relevant for the zonal travel cost method, check questions #50 and #52 are relevant for RUM studies, and the remaining questions are applicable to both types of travel cost studies. Questions #53–55 are about how successful the estimated model was from a statistical point of view.

Check questions	Yes/no/don't know	Comment
48. If the zonal travel cost method was used, were substitute sites taken into account?		
48a. If "yes", in what way?		
49. How was the travel cost model specified?		
50. If a RUM based study was used, how were site-specific data collected and quantified?		
51. Were individual-specific variables (i.e., variables whose values vary among respondents) used?		
52. How were respondents' choices of sites modelled?		
53. Was the total explanatory power (measured as, for example, adjusted R^2) significantly greater than zero?		
53a. If "yes", what was the explanatory power?		

Check questions	Yes/no/don't know	Comment
54. Was the confidence interval for the coefficient of the environmental quality variable reported?		
54a. If "yes", what was the interval?		
55. Was the confidence interval for the coefficient of the travel cost variable reported?		
55a. If "yes", what was the interval?		

3.3.4 CALCULATION OF TRAVEL COSTS

The travel cost is the sum of the expenditures needed for making the travel possible. It often consists of transportation costs, entrance fees, equipment costs and the opportunity cost of time. Costs for equipment that can be used also at other occasions and other costs that are not directly associated with the travel in question should not be included. A circumstance that might be difficult for the zonal travel cost method to handle is when the price of a given travel varies among respondents. For example, the price of air tickets for a particular flight might vary substantially among different groups of passengers. This tends to result in a weaker relation between the travel cost and the demanded number of travels, but this might be possible for the travel cost model to adjust for. The definition of the travel cost variable is crucial in a travel cost study and the estimates of recreational values are likely to show sensitivity for different definitions. A travel cost study of good quality has succeeded in making a reasonable definition. The following questions might provide guidance.

Check questions	Yes/no/don't know	Comment
56. What cost components were included in the travel cost?		
57. Did the study handle the problem that people's travel cost might vary for a given travel?		

3.3.5 OPPORTUNITY COST OF TIME

A travel cost study not considering the value of travel time might tend to underestimate the total travel cost and thus also recreational values. Most travel cost studies

use income as a basis for computing the value of travel time. A common assumption is that the value of time is some percentage of monthly wages/salaries (Cesario 1976). It is likely to be better to make such a crude assumption than not including the value of travel time at all. However, it should be noted that the travel might in some cases in itself increase the wellbeing of recreationists. Travel time involves a negative cost if this increase in wellbeing exceeds the opportunity cost of travel time. There is a risk for overestimation of recreational values if such a negative cost is not taken into account. Time spent on the recreational site (on-site time) should also be valued, because also this time could have been used for other activities, i.e. it has an opportunity cost. It is often assumed that on-site time is constant for all individuals and can be valued in the same way as travel time. It is common that the analyst estimates on-site time by looking at respondents' last travel to the site and the average time spent on the site at this occasion (see, e.g. Champ et al. 2003).

Estimating the opportunity cost of time is a part of a travel cost study's computation of total travel costs. This estimation is therefore important to consider when assessing the quality of the study. It can be noted that the value of travel time for an individual is not likely to be a constant, but depends on many different circumstances such as the purpose and length of the travel, the transport mode and if the travel takes place during the day or the night (SIKA 2002).

Check questions	Yes/no/don't know	Comment
58. Was the travel time valued?		
58a. If "yes", how was it valued?		
59. Does the study consider the possibility that the travel itself might contribute positively to the respondents' wellbeing?		
60. How was on-site time handled?		

3.3.6 MULTIPURPOSE TRIPS

Recreationists are often likely to have more than one purpose when visiting a recreational site. For example, a person travelling to a coastal area might be interested in bathing, fishing, walking as well as visiting restaurants. It might therefore be inadequate to allocate the whole travel cost to the activity linked to the environmental quality of interest to the analyst, e.g. fishing. Bathing and sun-bathing might have been a respondent's main purposes of visiting a site, even if she spent some time fishing. If so, allocating all travel costs to the fishing activity will overestimate the importance of fishing for this respondent. This suggests that it might be important in a travel cost study to ask respondents to report their purpose(s) of visiting a recreational site.

Check questions	Yes/no/don't know	Comment
61. If only one purpose of the travel was assumed, was there any motivation for this?		
62. If the travel had several purposes, was any corresponding adjustment of the travel costs made in the estimation of value estimates?		
63. Was there any discussion of the risk of overestimation/ underestimation of recreational values because of multipurpose trips?		

3.3.7 SELECTION OF ENVIRONMENTAL QUALITY VARIABLE

The selection of environmental quality is of crucial importance in a travel cost study, partly because such a variable makes it possible to link the recreational value for an improved environmental quality to measures for realising the improvement. Using RUM models is generally the most suitable approach for estimating recreational values caused by changes in environmental quality. Bockstael et al. (1987) describe what quantitative information has to be available for enabling an estimation of the recreational value of an abatement programme aiming at improving water quality. It would be desirable to have information on respondents' expectations about water quality and on how such expectations affected their choices of sites. In practice, it might be difficult to find data about expectations, which means that travel cost studies instead often use sample-specific data about realised choices or historical data. In general, it is an advantage if the environmental quality is assessed quantitatively by the visitors. For example, it is not very useful for an analyst to only know that the recreational fishing is "good". How should such a qualitative piece of information be interpreted and valued? Useful data would instead be, for example, fish catches (in weight or numbers) per effort. An alternative is to collect data about the environmental quality variable from "objective" measurements carried out by, for example, researchers, municipality officials, etc. However, it is in this case important that there is a connection between the "objectively" measured environmental quality and the visitors' perception of this quality.

It is also important to keep in mind that some effects of changes in environmental quality vary over time. The degree to which the effects and recreational data coincide in time might therefore be a factor that affects the results of the study.

Check questions	Yes/no/don't know	Comment
64. What environmental quality variable was used?		
65. Is the selected environmental quality variable relevant to the environmental problem in question?		
66. Is it possible for visitors to perceive quality measured by the selected environmental quality variable?		
67. Was the environmental quality assessed quantitatively by the visitors?		
68. If the environmental quality was measured "objectively", did the study investigate the relation between the objectively measured quality and visitors' perception of this quality?		
69. Did the study discuss to what extent the environmental effects and recreational data coincide in time?		

3.3.8 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.4 Quality factors for the property value method

A first step when carrying out a property value is typically to estimate a hedonic price function. Such a function describes how various housing and neighbourhood characteristics (attributes) influence the market price of a property. An estimated hedonic price function can subsequently be used for computing the so-called marginal implicit price of each attribute. A non-linear hedonic price function gives a (non-constant) relation between different attributes and the marginal implicit price. If an attribute related to environmental quality has been included, it might also be possible to estimate the willingness to pay for a change in this quality. The quality factors below were selected as particularly relevant for assessing the quality of a property value study:

3.4.1 Property values
3.4.2 Property attributes
3.4.3 Selection of environmental quality variable
3.4.4 Choice and estimation of model

3.4.1 PROPERTY VALUES

A property value study needs information on the dependent variable in the hedonic price function, i.e. the price variable. Price data from individual transactions at the property market are preferable, but property values assessed by tax authorities or property owners are sometimes used as price data. If assessed values are used, it is important to try to adjust for potential differences between such values and market prices. However, also market data might be biased. For example, market transactions between relatives or friends might not reflect the actual market value of a property. The price of renting a property is sometimes used in property value studies, which calls for cautiousness in interpreting the marginal implicit price, in particular if there are rent controls or other types of regulations.

Check questions	Yes/no/don't know	Comment
70. Did the study use market prices as data?		
71. If the study used non-market data, was the risk of measurement errors taken into account?		
71a. If "yes", in what way?		
72. Were data on individual market transactions used?		

Check questions	Yes/no/don't know	Comment
73. If more aggregated data (e.g., average prices in dwelling areas) were used, was there any discussion about the risk of errors in the measurement of prices because of this?		

3.4.2 PROPERTY ATTRIBUTES

Data on the property and its neighbourhood and environmental quality data are also necessary for a property value study valuing environmental change. A successful identification, definition and measurement of these attributes are important for results to be reliable. Examples of potentially important attributes include market transaction date (which might account for time trends in property values), distance to schools, parks and city centre, neighbourhood demographics such as average income, average age and ethnic composition, and environmental quality. Measurement errors and lack of data for important attributes are common problems in property value studies. It is still crucial that obviously important attributes are included in the analysis to the greatest extent possible.

Check question	Yes/no/don't know	Comment
74. What attributes were included in the analysis?		

3.4.3 SELECTION OF ENVIRONMENTAL QUALITY VARIABLE

The considerations in section 3.3.7 related to the selection of environmental quality variable in travel cost studies are valid also for property value studies. It is thus important for the property value method that the environmental quality attribute meets a number of criteria. The check questions are identical to those for the travel cost method, except for adjustments for the fact that the property value method is based on the behaviour of actors at the property market instead of the market for travels to recreational areas.

Check questions	Yes/no/don't know	Comment
75. What environmental quality variable was used?		

Check questions	Yes/no/don't know	Comment
76. Is the selected environmental quality variable relevant to the environmental problem in question?		
77. Is it possible for actors in the property market to perceive the quality measured by the selected environmental quality variable?		
78. Was the environmental quality assessed quantitatively by actors in the property market?		
79. If the environmental quality was measured "objectively", did the study investigate the relation between the objectively measured quality and actors' perception of this quality?		
80. Did the study discuss to what extent the environmental effects and property market data coincide in time?		

3.4.4 CHOICE AND ESTIMATION OF MODEL

It is important that a property value study make a reasonable definition of the property market, i.e. the population of properties subject to study. The delimitation of the market should neither be too wide nor too narrow. Further, the choice of functional form of the hedonic price function might have a substantial impact on results. It is seldom adequate to have a linear function, since this gives constant marginal implicit prices. Data structure and the analyst's judgments and expectations about the relation between attributes and property prices are in practice usually determining the choice of functional form. A challenge in a property value study is to go from the estimation of a hedonic price function to an estimation of a marginal WTP function. The marginal implicit price that can be computed from an estimated hedonic price function can only in special cases be interpreted as a marginal WTP. An estimation of a marginal WTP function might require a modelling of both the demand side and the supply side of the property market.

Check questions	Yes/no/don't know	Comment
81. How was the property market defined?		

Check questions	Yes/no/don't know	Comment
82. What was the functional form of the hedonic price function?		
82a. How was the choice of functional form motivated?		
83. Was the total explanatory power (measured as, for example, adjusted R^2) of the estimated hedonic price function significantly greater than zero?		
83a. If "yes", what was the explanatory power?		
84. Was the confidence interval for the coefficient of the environmental quality variable reported?		
84a. If "yes", what was the interval?		
85. Was there any test for multicollinearity?		
85a. If there was multicollinearity, what actions were taken because of this?		
86. Were the demand and supply sides of the property market modelled?		
87. Besides estimation of the marginal implicit price, was the willingness to pay for the environmental change estimated?		

3.4.5 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.5 Quality factors for the defensive expenditure method

The following quality factors were identified for the defensive expenditure method:

3.5.1 Properties of the good
3.5.2 Procedure for estimation of the economic value

3.5.1 PROPERTIES OF THE GOOD

The point of departure for the defensive expenditure method is to find a market good serving as a substitute for an ecosystem service/environmental quality. One example of such a market good is water filters giving protection against a reduced drinking water quality. However, the studied market good might differ in several respects from the ecosystem service, which suggests that a willingness to pay for the market good cannot be interpreted straight off as a willingness to pay for the ecosystem service. For example, the market good might be far from a perfect substitute, it might provide other types of utility than the ecosystem service, and it might cause negative side effects. It is thus necessary to analyse to what extent the market good have properties that differ from those of the ecosystem service.

The degree of substitutability affects the choice of valuation procedure. If the market good and the ecosystem service are perfect substitutes, a small change in the provision of the ecosystem service can be valued as the change in expenditures for the market good. However, it is more complicated to value non-marginal changes in the provision of the ecosystem service. For example, if there is a substantial improvement in drinking water quality, an individual might decrease the consumption of water filters and still enjoy a given utility level. It is then necessary to consider the fact that the individual can use this situation for changing her consumption levels also of other goods than water filters. This means that the decrease in expenditures for water filters constitutes a lower boundary for the economic value of an improved drinking water quality (Freeman 2003). Estimations of economic values become more complicated in cases when it is less reasonable to view the ecosystem service and the market good as perfect substitutes.

An important aspect to consider is that the individuals' perceived protection from consuming the market good might differ from the protection that is scientifically established. Health risk literature makes a distinction between subjective risk and objective risk. A difference between subjective and objective risk might make it difficult to relate an estimated willingness to pay to the protection level that objectively is associated with consumption of the good.

Check questions	Yes/no/don't know	Comment
88. The market good might have other properties than the valued ecosystem service/ environmental quality. Was there an analysis of such potential differences?		
88a. If "yes", how were the market good's potential weaknesses as a substitute taken into account?		
88b. If "yes", how was the possibility that the consumption of the market good might result in other types of utility than those provided by the ecosystem service taken into account?		
88c. If "yes", how was the possibility that the consumption of the market good potentially causes side effects giving disutility taken into account?		
89. How were potential differences between the subjectively perceived level of protection and the objectively determined level of protection, caused by the market good, taken into account?		

3.5.2 PROCEDURE FOR ESTIMATION OF THE ECONOMIC VALUE

Defensive expenditure studies are characterised by very different levels of ambition concerning the collection and use of economic data. The simplest studies only use data on the market good's protection capacity and data on prices or expenditures. Such information might be sufficient for some applications, for example, in cases when the market good and the ecosystem service are close substitutes and a small change in the provision of the ecosystem service is to be valued. More advanced defensive expenditure studies also estimate a demand function for the market good and/or a health production function (Dickie 2003). It is in general an advantage to carry out such estimations. For example, an

estimated demand function makes it possible (for a knowledgeable analyst) to estimate changes in consumer surplus caused by non-marginal changes in the provision of the ecosystem service.

Check questions	Yes/no/don't know	Comment
90. Was a demand function for the market good and/or a health production function estimated?		
90a. If "yes", how was the function specified and what was its explanatory power?		
90b. If "no", what procedure was followed for the estimation of economic values and how was it motivated?		

3.5.3 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

3.6 Quality factors for stated preferences methods

The following quality factors were identified for stated preferences methods. The last two factors are specific for two main stated preferences methods: the contingent valuation method and choice experiments.

3.6.1 Acceptance and understanding of the valuation scenario
3.6.2 Description of effects of the environmental change
3.6.3 Information on the null alternative
3.6.4 Winners or losers?
3.6.5 Payment and delivery conditions
3.6.6 Willingness to pay or willingness to accept compensation?
3.6.7 Valuation function
3.6.8 Test for hypothetical bias
3.6.9 Specific quality factors for the contingent valuation method (CVM)
3.6.10 Specific quality factors for choice experiments (CE)

3.6.1 ACCEPTANCE AND UNDERSTANDING OF THE VALUATION SCENARIO

It is crucial in a stated preferences study that respondents understand and accept the valuation scenario, and also grasp the information about the environmental change included in the survey instrument. A basic requirement is therefore that the valuation study reports the valuation scenario and the valuation questions. However, also the other parts of the survey instrument should be reported, since also questions and information in these other parts might influence the respondents' interpretation of the valuation scenario. Check question #35 should thus have been answered in the affirmative.

It might be very difficult to assess the degree of acceptance and understanding among respondents, unless this was subject to a special analysis in the valuation study. Such an analysis can be based on, for example, respondents' answers to follow-up questions that in various ways test for understanding and acceptance. High item non-response for the valuation questions and many protests against the valuation scenario might also indicate an unsatisfactory valuation scenario. Questions such as "why didn't you answer the valuation question?" might be suitable for identifying protest answers (cf. Jorgensen et al. 1999, Söderqvist 1998).

An insufficient sensitivity of value estimates for the size of the environmental change might also be an indication of weaknesses in the valuation scenario. What is insufficient or not is typically difficult to say in advance, but it is reasonable to question the valuation scenario if a so-called scope test shows that the value of a very small change is not significantly different from the value of a very big change.

Check questions	Yes/no/don't know	Comment
91. Did the valuation study include an analysis of the extent to which respondents understood and accepted the valuation scenario (including information on the environmental change)?		
91a. If "yes", what was the result of the analysis?		
92. What was the item non-response rate for the valuation question(s) (in percent)?		
93. Were protest answers identified?		
93a. If "yes", how common were protest answers? (For example, what proportion of the item non-response for the valuation question can be attributed to protests?)		
93b. If "yes", in what way were the protest answers taken into account in the estimation of economic values?		
94. Was any scope test carried out?		
94a. If "yes", what was the result of the test?		

3.6.2 DESCRIPTION OF EFFECTS OF THE ENVIRONMENTAL CHANGE

It must be clear from the valuation scenario how people are directly or indirectly affected by the environmental change. That is, the effects of the environmental change must be described, and this has to be done as objectively as possible. However, how this description can be designed varies from case to case. It is in some cases possible to make a quantitative description (e.g. a reduction in a particular death risk from 200 to 2 Swedes per year, or a reduction in an average noise level from 70 dB to 50 dB), and in other cases the description will be more qualitative (e.g. a description of the effects of the environmental change by drawings or photographs). However, the effects must somehow be described; it is not reasonable to ask the respondents to value, for example, a 1000 tonnes

reduction of the emissions of some substance if no information is given about how this reduction affects environmental quality, the provision of ecosystem services, etc.

Check questions	Yes/no/don't know	Comment
95. Were the *effects* of the environmental change described to the respondents?		
96. Can the effects in the valuation scenario be deemed to be objectively described?		

3.6.3 INFORMATION ON THE NULL ALTERNATIVE

It is not a reasonable task for the respondents to value a project that would result in some environmental change without being informed about what will happen if the project is not realised. Examples of such information are that the environmental quality will remain at today's level or that the environmental quality will be reduced to a certain level. The effects of such an alternative scenario ("the null alternative") must also be described in the valuation scenario.

Check question	Yes/no/don't know	Comment
97. Were the effects of a null alternative described in the valuation scenario?		

3.6.4 WINNERS OR LOSERS?

A realised valuation scenario might increase or decrease a respondent's well-being, or not affect her wellbeing at all. It is an advantage if a valuation study captures all these three groups, and thus avoids unintentional delimitations to, for example, only those respondents who gain from a realised scenario.

A related problem is that the results of the valuation study might be difficult to interpret if its scenario implies both advantages and disadvantages to some respondents. For example, if the effects of a nutrient abatement programme are to be valued by those who is supposed to contribute to the abatement (e.g. farmers), they might subtract their expected abatement costs from their benefits of a reduced eutrophication. In such a case, it should be clear from the valuation scenario if the respondents are asked to base their answers to the valuation questions on their gross benefits (before deduction for costs) or on their net benefits (after deduction for costs).

Check questions	Yes/no/don't know	Comment
98. Does the valuation study capture not only those gaining from a realised valuation scenario, but also those losing and those whose wellbeing are not affected?		
99. Did the valuation scenario make clear how respondents should take into account their potential costs for a realization of the scenario?		

3.6.5 PAYMENT AND DELIVERY CONDITIONS

The valuation scenario should describe the conditions determining whether the project would be carried out or not. One example of such a delivery condition is that the project is realised if the total benefits exceed the total costs associated with the project. This is a natural condition from a cost-benefit analysis point of view, but other types of delivery conditions are also conceivable. One example of an alternative condition is that the project is realised if a majority of respondents vote "yes" to the project.

How respondents would pay for the project if it is realised is another important piece of information. Such payment conditions include the type of payment vehicle (e.g. an ear-marked environmental tax, water fees, contributions to a fund, etc.), the time of payment (e.g. a once-for-all amount, a monthly payment during one year, an annual payment during ten years, etc.), and how the size of the respondent's payment is determined. The payment might be described as proportional to or equal to the willingness to pay reported by the respondent, but other payment conditions are sometimes used. For example, fairness considerations might suggest that a respondent's payment should be related to her income. The design of the payment conditions might have a great impact on the number of protest answers.

Check questions	Yes/no/don't know	Comment
100. Were delivery conditions specified?		
100a. If "yes", how were they specified?		

Check questions	Yes/no/don't know	Comment
101. Was there a specification of the determinants of the size of each respondent's payment in the case of a realised project?		
101a. If "yes", how were these determinants specified?		
102. What payment vehicle was used?		
103. If the payment was not a once-for-all amount, was it specified how often and for what length of time the payments would be made?		

3.6.6 WILLINGNESS TO PAY OR WILLINGNESS TO ACCEPT COMPENSATION?

Economic theory indicates the contexts in which valuation questions should be eliciting willingness to pay or willingness to accept compensation, see section 2.3. However, WTA questions are empirically problematic (cf. Horowitz and McConnell 2002). They tend to cause a substantial number of protest answers. Further, it is not uncommon that some respondents give very high WTA amounts as answers to open-ended WTA questions. This can be viewed as a kind of protest, but it might also reflect that WTA (in contrast to WTP) is not limited by any budget restriction. The empirical problems associated with WTA questions suggest that questions eliciting WTP are a better choice, given that the valuation context does not suggest special reasons for preferring questions eliciting WTA.

Check questions	Yes/no/don't know	Comment
104. Was the aim of the valuation question to obtain information about willingness to pay or about willingness to accept compensation?		
104a. If the aim was to obtain information about willingness to accept compensation, what were the reasons for this choice?		

3.6.7 VALUATION FUNCTION

A stated preferences study often involves the estimation of a so-called valuation function as a way of relating the respondents' answers to the valuation question to various explanatory variables, such as their answers to socio-economic questions and other characteristics of the respondents and the environmental change subject to valuation. The shape of the valuation function depends on, inter alia, the framing of the valuation question. Two examples: (i) if the respondents were free to state WTP amounts, the function might have been estimated by a simple regression analysis with WTP as the dependent variable; (ii) if the respondents were asked to answer "yes" or "no" to a given monetary amount (a "bid"), the probability of a "yes" answer is usually the dependent variable and the bid is one of the explanatory variables in a discrete choice model such as the logit model.

There are sometimes reasons to expect a positive or negative sign of the coefficients of the explanatory variables. For example, income and WTP are likely to be positively correlated (cf. Hökby and Söderqvist 2003). An estimated valuation function thus provides an opportunity to evaluate whether the valuation results are reasonable or not. It might also be useful in attempts to generalise the results of the valuation study to other settings. However, a common problem is a low total explanatory power (measured as, e.g. adjusted R^2). However, while it is desirable that it should be statistically greater than zero (i.e. a null hypothesis that all coefficients of the explanatory variables are equal to zero can be rejected at a level of significance ≤ 10%), it is left to the analyst to judge what level above zero is a minimum acceptable explanatory power.

Check questions	Yes/no/don't know	Comment
105. Was any valuation function estimated?		
105a. If "yes", was the total explanatory power significantly greater than zero?		
105b. If "yes", what was the total explanatory power?		
105c. If "yes", to what extent were the signs of the coefficients the expected ones?		
105d. If the valuation function included an income variable, was the coefficient of this variable positive?		

3.6.8 TEST FOR HYPOTHETICAL BIAS

Stated preferences methods are characterised by the fact that actual market transactions do not take place, and this hypothetical nature of the valuation has been heavily criticised. It is therefore an advantage if the valuation study tried to adjust for the consequences that the hypothetical setting might cause. There are some adjustment methods available. For example, the respondents could be asked to state the degree of certainty associated with their answers to the valuation question (Champ and Bishop 2001) or to answer questions that make them to discuss why there might be a hypothetical bias, so-called "cheap-talk script" (Cummings and Taylor 1999).

A conspicuously high willingness to pay in relation to respondents' income might be interpreted as a sign of hypothetical bias if there are indications that respondents have not adequately taken their budget constraint into account. Alternative interpretations are that this indicates protests against the valuation scenario, or simply unusually strong preferences for a realisation of the scenario.

Check questions	Yes/no/don't know	Comment
106. Was any adjustment made for hypothetical bias?		
106a. If "yes", what kind of adjustment?		
107. Was the size of respondents' willingness to pay studied in terms of its proportion of their income?		
107a. If "yes", what was the average proportion?		

3.6.9 SPECIFIC QUALITY FACTORS FOR THE CONTINGENT VALUATION METHOD

There is an extensive discussion on the framing of willingness to pay questions in CVM studies, and it has been shown that the choice of elicitation format influences estimates of, for example, mean WTP. Closed-ended questions giving yes/no answers minimise the risk of strategic behaviour and resemble a normal market situation, but they give scarce information on the respondent's willingness to pay. Moreover, there are advices against the use of closed-ended questions about attitudes, opinions and values because respondents often show a tendency to agree (yea-saying bias) (SCB 2001b). This phenomenon might be an important reason to why mean WTP estimates based on answers to closed-ended questions tend to be greater than such estimates based on answers to open-ended questions

(Kriström 1993). Table 11 reports the main elicitation formats used by CVM studies and what effects the formats usually are believed to cause. What elicitation format should be preferred is, in our opinion, an unresolved issue. However, it is important that the elicitation format is reported, since this might indicate how results should be interpreted. It is also an advantage if the study analysed the effects of using different elicitation formats. This might give important information on the sensitivity of valuation results to the elicitation format.

A special problem associated with questions giving respondents an opportunity to state their WTP is that this might result in some very high WTP amounts. This introduces a difficulty to judge whether such WTP amounts are reasonable or should be interpreted as protests. The treatment of outliers might influence mean WTP estimates substantially, and the valuation study should therefore report clearly if outliers were excluded from the analysis and, if so, the criteria for exclusion.

All elicitation formats except purely open-ended questions involve monetary amounts that the respondent is asked to consider. Some formats allow each respondent to choose among several different amounts. Other formats only involve one amount followed by a question on whether the respondent is willing or not willing to pay the amount. The amount (the ”bid”) is varied among different respondents, which means that respondents' yes/no answers together give information on the WTP distribution. It is an advantage if the design of such a ”bid vector” results in a large proportion of ”yes” answers among those respondents who meet the lowest bid and large proportion of ”no” answers among those respondents who meet the highest bid. The estimates of, for example, mean WTP might otherwise be associated with a high degree of uncertainty. To include monetary amounts in the valuation question might give rise to anchoring effects in the sense that respondents use the amounts as indications of what willingness to pay they should have. It might therefore be important to analyse if there are any anchoring effects or not. For example, if payment cards are used, such a test can be carried out if at least two different designs of the cards were used in the study.

Table 11. *Main elicitation formats in CVM studies and some stylised facts.*

Open-ended	Large number of zero responses, few small positive responses.
Bidding game	Final estimate shows dependence on starting point used.
Payment card	Weak dependence of estimate on amounts used in the card.
Single-bounded dichotomous choice	Population WTP estimates typically higher than other formats.
Double-bounded dichotomous choice	The two responses do not correspond to the same underlying WTP distribution.

Source: Bateman et al. (2002).

Check questions	Yes/no/don't know	Comment
108. What type of elicitation format was used?		
109. Were the consequences of using different types of elicitation formats tested?		
109a. If "yes", what was the result of the test?		
110. Did the study report how WTP outliers (if any) were treated?		
110a. If "yes", how were they treated in the analysis?		
111. If a bid vector was used, what proportion of the respondents accepted the *lowest* bid?		
112. If a bid vector was used, what proportion of the respondents accepted the *highest* bid?		
113. Was there an analysis of the presence of potential anchoring effects?		
113a. If "yes", what was the result of the analysis?		

3.6.10 SPECIFIC QUALITY FACTORS FOR CHOICE EXPERIMENTS

A special feature of CE studies is that the environmental change is described by a number of attributes ("environmental attributes") and that a willingness to pay can be estimated for changes in individual attributes. Information that makes such an estimation possible is obtained by varying the levels of the environmental attributes and the cost of realising the environmental change ("the cost attribute") among the respondents.

What attributes are suitable to choose vary from case to case. In any case, the respondents need clear definitions of the attributes. The choice situation tends to become very complicated to the respondents if the number of attributes is large. What is a "large number" depends on the context, but it can be observed that the number of attributes usually does not exceed 4-7. A case where two or several attributes are mutually dependent might also make the choice situ-

ation difficult to the respondents. A strong dependence of this kind has to be taken into account by the CE study. It is not only the number of attributes that determine the degree of complexity of the task that the respondent is asked to perform. This degree is also dependent on the number of alternatives that the respondent is asked to choose among in each choice situation, and the number of choice situations that the respondent meet. A too demanding task might result in a substantial item non-response for the choice situations, cf. check question #92.

The use of several attributes which each can take several different values implies a very large number of possible combinations. CE studies therefore usually employ various methods for reducing the number of combinations in order to efficiently design the choice set, cf. Alpízar et al. (2003) and Bateman et al. (2002). The study should report what design technique was used.

The inclusion of a null alternative as one of the alternatives that the respondents can choose among usually simplifies the estimation of valuation estimates or at least respondents' understanding for the alternatives available. It might thus be an advantage for a CE study to not only include a description of the null alternative (cf. check question #97), but also make it possible for the respondents to choose this alternative.

It is an advantage if the CE study involved tests for internal validity. For example, such tests might show if respondents' answers are consistent with desirable properties of their preferences, such as transitivity. See Alpízar et al. (2003).

Check questions	Yes/no/don't know	Comment
114. How many attributes were used in total (including the cost attribute)?		
115. Did the study account for potential dependence among the environmental attributes?		
116. How many alternatives did the respondent choose from in each choice situation?		
117. How many choice situations did the respondent face?		
118. What technique was used to design the choice set?		

Check questions	Yes/no/don't know	Comment
119. Was it possible for the respondent to choose a null alternative?		
120. Was there any test for internal validity of the respondents' answers?		
120a. If "yes", what was the result of the test?		

3.6.11 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.7 Quality factors for the replacement cost method

Appendix A reports three general conditions for the validity of the replacement cost method as a valuation method, cf. Shabman and Batie (1978), Bockstael et al. (2000) and Freeman (2003). To assess the quality of a replacement cost study is therefore much about judging to what extent these conditions are fulfilled. It can be noted that a literature review by Sundberg (2004) showed that some replacement cost studies do not even discuss whether the conditions can be regarded as fulfilled or not. The conditions can be expressed as the following quality factors:

3.7.1 The performance of the man-made system as a substitute
3.7.2 The cost-effectiveness of the man-made system
3.7.3 Willingness to pay for replacement costs?

3.7.1 THE PERFORMANCE OF THE MAN-MADE SYSTEM AS A SUBSTITUTE

The background to the replacement cost method is that humans might in some cases be able to replace the loss of an ecosystem service by, for example, some technological solution. However, to introduce a man-made substitute implies costs, and society would not have to pay these costs if the ecosystem service is available. ”Replacement” suggests that the man-made system should be able to provide services of similar quantity and quality as those provided by the ecosystem. If the man-made system is a poor substitute for the ecosystem in the sense that there are great differences in quantity and quality, the replacement cost method might result in substantially biased estimates of the economic value of the ecosystem service.

Check questions	Yes/no/don't know	Comment
121. Did the study analyse if the man-made system provides a service of the same quantity and quality as the ecosystem service subject to valuation?		
121a. If ”yes”, how was this analysis carried out?		
122. If there is any important difference in quantity or quality, was this taken into account in the esti-mation of values?		

3.7.2 THE COST-EFFECTIVENESS OF THE MAN-MADE SYSTEM

The replacement cost method is likely to overestimate the value of the ecosystem service if the man-made system is replacing the ecosystem service in an unnecessarily costly way. If there are alternative man-made systems or alternative designs of a particular man-made system, the replacement cost study should identify the system or the design that replaces the ecosystem service at the lowest possible costs.

It might be a difficult empirical task to include all relevant fixed and variable costs associated to the man-made system. The relevant types of fixed and variable costs vary from case to case, but it might be particularly important to check that the study has not ignored potential initial investment costs for the man-made system.

Check questions	Yes/no/don't know	Comment
123. Did the study analyse the possibility that the ecosystem service might be replaced in several different ways?		
124. Was the valuation based on the cost-effective way of replacing the ecosystem service?		
125. Did the study report what types of fixed and variable costs constitute the basis of the valuation?		
126. Did the cost estimation take potential initial investment costs into account?		

3.7.3 WILLINGNESS TO PAY FOR REPLACEMENT COSTS?

To carry out economic valuation by using data on replacement costs is risky from a welfare economics point of view because the costs are not necessarily covered by citizens' willingness to pay. Information on willingness to pay is obtained by applying RP or SP methods (cf. appendix A), but such information does probably not exist if a replacement cost study was carried out. This is simply because the existence of an RP or SP study would make the replacement cost study unnecessary from a benefit estimation perspective. (However, a replace-ment cost study could very well be needed for obtaining cost estimates to be compared to benefits in a cost-benefit analysis.) Even if any explicit WTP

information is thus not likely to be available when a replacement cost study is carried out for valuation purposes, the study should discuss if there are indications of presence of a willingness to pay for the replacement costs.

Check questions	Yes/no/don't know	Comment
127. Were there indications that individuals would be willing to pay the replacement costs if the ecosystem service was not available?		

3.7.4 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.8 Quality factors for the human capital method

The following quality factors were identified for the human capital method:

3.8.1 Theoretical considerations
3.8.2 Technological development
3.8.3 To estimate the value of lost productivity

A study applying the human capital method is likely to benefit from advice from medical experts. This suggests that it is an advantage if check question #39 was answered in the affirmative.

3.8.1 THEORETICAL CONSIDERATIONS

Maybe the most important theoretical consideration to make in a human capital study is to judge if and how the results can be interpreted in terms of willingness to pay. Costs of medical care are not a measure of willingness to pay for avoiding illness but rather a cost of illness *ex post*. The human capital method is based on two main assumptions: (1) direct costs of illness reflect the economic value of goods and services used for treating the illness, and (2) an individual's income reflects the economic value of lost production. The method thus measure costs *ex post* and does not attempt to measure reductions in wellbeing due to illness. Further, costs *ex post* do not reflect any variability in individuals' risk attitudes. Despite these limitations of the human capital method, estimates from applications of the method are still often interpreted as minimum estimates of willingness to pay.

Check questions	Yes/no/don't know	Comment
128. Were the results of the human capital study interpreted in terms of willingness to pay?		
128a. If "yes", how was this interpretation made?		

3.8.2 TECHNOLOGICAL DEVELOPMENT

A human capital study should consider the development of methods and techniques for medical treatments. Technological change might influence the cost of illness and thus increase or decrease value estimates for a given illness. Before using results from human capital studies, it is important to find out if and how treatment methods have changed after the study was made.

Check questions	Yes/no/don't know	Comment
129. If there have been substantial changes in treatment methods relevant to the study, did the study take such changes into account?		
129a. If "yes", was it discussed how these changes influence value estimates?		

3.8.3 TO ESTIMATE THE VALUE OF LOST PRODUCTIVITY

There are different ways of measuring losses in working time. To only estimate the reduced hours of work due to illness is not likely to reflect the total loss of an individual's production correctly. This is particularly true in the case of prolonged chronic illness. Such illness might force an individual to work part-time, to accept a lower wage, or even to leave the labour market. As regards the value of lost time, its estimation should include the productivity of individuals not having paid jobs, and preferably also the value of lost leisure time.[1]

Check questions	Yes/no/don't know	Comment
130. How were losses in working time estimated?		
131. Did the study take chronic illness into account?		
132. How was the cost of time defined?		
133. Did the estimated value of lost time include the productivity of individuals who work outside the labour market?		

[1] The discussion about so-called *friction costs* could be noted in this context. It has been argued that the presence of unemployment makes it rather simple to replace individuals who suffer from illness by unemployed individuals (e.g. Koopmanschap et al. 1995). However, friction costs arise because the replacement process might take some time. Accepting this argument and limiting the analysis to such costs would imply other results than those from applications of the human capital method, which do not ignore the value of individuals' potential production even if they are replaced. For a critique of the friction cost approach, see Johannesson and Karlsson (1997).

Check questions	Yes/no/don't know	Comment
134. Did the estimated value of lost time also include the value of lost leisure time?		

3.8.4 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.9 Quality factors for valuation based on the costs of realising political decisions

This method is similar to the replacement cost method (see section 3.7) in the sense that it makes use of cost data for valuing environmental change. For example, the data might be about the costs of improving or restoring the environment. The following two quality factors were identified:

3.9.1 Cost-effectiveness
3.9.2 Willingness to pay for the costs?

3.9.1 COST-EFFECTIVENESS

There are usually several different ways of implementing a political decision, and it is important that it is analysed what way is the least-cost alternative.

It might be a difficult empirical task to include all relevant fixed and variable costs associated with the implementation of the decision. What types of fixed and variable costs are relevant vary from case to case, but it might be particularly important to check that the study has not ignored potential initial investment costs for the implementation.

Check questions	Yes/no/don't know	Comment
135. Did the study analyse the possibility that the political decision might be realised in several different ways?		
136. Was the valuation based on the cost-effective way of realising the decision?		
137. Did the study report what types of fixed and variable costs constitute the basis of the valuation?		
138. Did the cost estimation take potential initial investment costs into account?		

3.9.2 WILLINGNESS TO PAY FOR THE COSTS?

To carry out economic valuation by using cost data is risky from a welfare economics point of view because the costs are not necessarily covered by citizens' willingness to pay. Information on willingness to pay is obtained by applying RP or SP methods (cf. appendix A), but such information does probably not exist if there is an attempt to make valuation by studying costs. This is simply because the existence of an RP or SP study would make the cost study unnecessary from a benefit estimation perspective. (However, the cost study could very well be needed for obtaining cost estimates to be compared to benefits in a cost-benefit analysis.) Even if any explicit WTP information is thus not likely to be available when a cost study is carried out for valuation purposes, the study should discuss if there are indications of presence of a willingness to pay for the costs.

Check question	Yes/no/don't know	Comment
139. Were there indications that the estimated costs could be covered by citizens' willingness to pay?		

3.9.3 SUPPLEMENTARY COMMENTS

Supplementary comments, e.g. on whether "no"/"don't know" answers indicate low quality or not:

_____

_____

_____

3.10 Overall quality assessment

This final part of the quality assessment instrument gives an opportunity to make an overall assessment of the quality of the valuation study. The overall assessment should consider both quality factors applicable to all valuation studies (section 3.1) and specific quality factors for valuation methods (sections 3.2–3.9). It is important in the overall assessment to avoid a myopic perspective but instead consider broader issues such as:

- What was the valuation study aiming at estimating?
- How was the population defined?
- To what extent did the valuation study succeed in measuring what it aimed at measuring for the population in question?
- Were there any risks of double counting of economic values? If so, was double counting avoided? For example, there is a risk of double-counting if a CVM study is aiming at valuing the establishment of two different nature reserves and does not clarify whether the respondent should value the establishment of nature reserve #1 given that nature reserve #2 has been established or given that nature reserve #2 has not been established.

Overall quality assessment:

References

Alpízar, F., Carlsson, F., Martinsson, P., 2003. Using choice experiments for non-market valuation. Economic Issues 8, 83-110.

Andrews, T., 1996. A discrete choice model of recreational trout angler benefits in Pennsylvania. Unpublished manuscript, Department of Economics, West Chester University.

Arrow, K., Cline, W. R., Mäler, K-G., Munasinghe, M., Squitieri, R., Stiglitz, J., 1996. Intertemporal equity, discounting and economic efficiency, pp. 125-144 in Bruce, J. P., Lee, H., Haites, E. F., Climate Change 1995: Economic and Social Dimensions of Climate Change. Cambridge University Press, Cambridge, UK.

Arrow, K., Solow, R., Leamer, E., Portney, P., Randner, R., Schuman, H., 1993. Report of the NOAA Panel on Contingent Valuation. Federal Register 58(10), 4601-4614.

Bann, C., Fisher, J., Horton, B., 2003. The Benefits Assessment Guidance for PR04: Review of Non-use Values for Water Quality and Water Resources and Values for Bathing Water Improvements. Report of an Expert Workshop in Peterborough, May 2003. UK Environment Agency. Available at http://www.environment agency.gov.uk/commondata/103599/petreport_673271.doc (15 November 2004).

Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, D. W., Sugden, R., Swanson, J., 2002. Economic Valuation with Stated Preferences Techniques: A Manual. Edward Elgar Publishing, Cheltenham, UK.

Biemer, P. P., Lyberg, L. E., 2003. Introduction to Survey Quality. John Wiley & Sons, Hoboken, NJ.

Bishop, R. C., 2003. Where to from here? In: Champ, P. A., Boyle, K. J., Brown, T. C. (eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 537-566.

Bockstael, N. E., Hanemann, W. M., Strand, I. E. (eds.), 1987. Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models. Report to U.S. Environmental Protection Agency, EPA Contract No. CR-811043-01-0.

Bockstael, N. E., Freeman, A. M., Kopp, R. J., Portney, P. R., Smith, V. K., 2000. On measuring economic values for nature. Environmental Science and Technology 34, 1384-1389.

Brookshire, D. S., Ives, B. C., Schulze, W. D., 1976. The valuation of aesthetic preferences. Journal of Environmental Economics and Management 3, 325-346.

Carson, R. T., 2000. Contingent valuation: a user's guide. Environmental Science and Technology 34, 1413-1418.

Carson, R. T., in preparation. Contingent Valuation: A Comprehensive Bibliography and History. Edward Elgar Publishing, Cheltenham, UK.

Carson, R. T., Flores, N. E., Meade, N. F., 2001. Contingent valuation: controversies and evidence. Environmental and Resource Economics 19, 173-210.

Carson, R. T., Mitchell, R. C., Hanemann, W. M., Kopp, R. J., Presser, S., Ruud, P. A., 1992. A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. A Report to the Attorney General of the State of Alaska. Available at http://www.evostc.state.ak.us/pdf/econ5.pdf (9 November 2004).

Champ, P. A., Bishop, R., 2001. Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. Environmental and Resource Economics 19, 383-402.

Champ, P. A., Boyle, K. J., Brown, T. C., 2003. A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht, The Netherlands.

CIE, Centre for International Economics, 2001. Review of willingness-to-pay methodologies. Report prepared for Independent Pricing and Regulatory Tribunal of New South Wales. CIE, Canberra. Available at http://www.ipart.nsw.gov.au/pdf/cie.pdf (11 October 2004).

Cochran, W. G., 1977. Sampling Techniques, Third Edition. Wiley, New York.

Cummings, R. G., Taylor, L. O., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. American Economic Review 89, 649-665.
Daily, G. C., Söderqvist, T., Aniyar, S., Arrow, K., Dasgupta, P., Ehrlich, P. R., Folke, C., Jansson, A., Jansson, B-O., Kautsky, N., Levin, S., Lubchenco, J., Mäler, K-G., Simpson, D., Starrett, D., Tilman, D., Walker, B., 2000. The value of nature and the nature of value. Science 289, 395-396.
Dalenius, T., 1985. Elements of Survey Sampling. Swedish Agency for Research Cooperation with Developing Countries (SAREC), Stockholm.
Davis, D. D, Holt, C. A., 1993. Experimental Economics. Princeton University Press, Princeton, NJ.
Defra (Department of Environment, Food and Rural Affairs), 2000. Flood and Coastal Defence Project Appraisal Guidance: Economic Appraisal. A Procedural Guide for Operating Authorities. Available at http://www.defra.gov.uk/environ/fcd/pubs/pagn/fcdpag3 (11 October 2004).
Dickie, M., 2003. Defensive behavior and damage cost methods, pp. 395-444 in Champ, P. A., Boyle, K. J., Brown, T. C. (eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Dillman, D., 1978. Mail and Telephone Surveys: The Total Design Method. Wiley-Interscience, New York.
Dillman, D., 1991. The design and administration of mail surveys. Annual Review of Sociology 17, 225-249.
Dillman, D., 2000. Mail and Internet Surveys: The Tailored Design Method. Wiley-Interscience, New York.
Freeman III, A. M., 2003. The Measurement of Environmental and Resource Values: Theory and Methods. Second Edition. Resources for the Future, Washington, DC.
Frykblom, P., Helgesson, A., 2002. Cost-benefitanalys inom miljöområdet: en kartläggning. Rapport 5198, Naturvårdsverket, Stockholm.
Haab, T. C., McConnell, K. E., 2002. Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation. Edward Elgar Publishing, Cheltenham, UK.
Hausman, J. A. (ed.), 1993. Contingent Valuation: A Critical Assessment. North-Holland, Amsterdam.
Horowitz, J., McConnell, K. E., 2002. A review of WTA/WTP studies. Journal of Environmental Economics and Management 44, 426-447.
Hökby, S., Söderqvist, T., 2003. Elasticities of demand and willingness to pay for environmental services in Sweden. Environmental and Resource Economics 26, 361-383.
Japec, L., Ahtiainen, A., Hörngren, J., Lindén, H., Lyberg, L., Nilsson, P., 1997. Minska bortfallet. Current Best Methods-rapport. Statistiska Centralbyrån, Örebro.
Johannesson, M., Karlsson, G., 1997. The friction cost method: A comment. Journal of Health Economics 16, 249-255.
Jorgensen, B. S., Syme, G. J., Bishop, B. J., Nancarrow, B. E., 1999. Protest responses in contingent valuation. Environmental and Resource Economics 14, 131-150.
Juran, J. M., Gryna, Jr., F. M., 1980. Quality Planning and Analysis, Second Edition. McGraw-Hill, New York.
Kerr, G., Lincoln University. E-mail message, 25 October 2004.
Koopmanschap, M. A., Rutten, F. F. H., van Ineveld, B. M., van Roijen, L., 1995. The friction cost method for measuring indirect costs of disease. Journal of Health Economics 14, 675-682.
Kriström, B., 1992. Hur mycket vill svenska folket betala för en bättre miljö? Ekonomisk Debatt 3, 209-219.
Kriström, B., 1993. Comparing continuous and discrete contingent valuation questions. Environmental and Resource Economics 3, 63-71.
Kriström, B., 1996. Contingent valuation. Arbetsrapport 233, Institutionen för skogsekonomi, Sveriges Lantbruksuniversitet, Umeå.
Lundström, S., Särndal, C-E., 2001. Estimation in the Presence of Nonresponse and Frame Imperfections. Current Best Methods-rapport. Statistiska Centralbyrån, Örebro.

McComb, G., Economist, Environment Canada. E-mail message, 20 August 2004.
Mitchell, R. C., Carson, R. T., 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.
Nash, K., Assistant Economist, New South Wales Environmental Protection Agency. E-mail message, 24 August 2004.
Naturvårdsverket 2004. Ekonomiska konsekvensanalyser i myndigheternas miljöarbete – förslag till förbättringar. Rapport 5398, Naturvårdsverket, Stockholm.
Parsons, G. R., Kealy, M. J., 1992. Randomly drawn opportunity sets in a random utility model of recreation. Land Economics 68 (4), 418-433.
Phaneuf, D. J., Smith, V. K., 2004. Recreation demand models. In: Mäler, K-G., Vincent, J. (eds.), Handbook of Environmental Economics. North-Holland, Amsterdam.
Randall, A., Ives, B., Eastman, C., 1974. Bidding games for valuation of aesthetic environmental improvements. Journal of Environmental Economics and Management 1, 132-149.
Sagoff, M., 1988. The Economy of the Earth: Philosophy, Law, and the Environment. Cambridge University Press, Cambridge, UK.
Samuelson, P., 1954. The pure theory of public expenditure. Review of Economics and Statistics 36, 387-389.
SCB, 2001a. Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration i officiell statistik. Meddelande i samordningsfrågor (MIS), Statistiska Centralbyrån, Stockholm.
SCB, 2001b. Fråga rätt! Utveckla, testa, utvärdera och förbättra blanketter. Current Best Methods-rapport. Statistiska Centralbyrån, Stockholm.
Shabman, L. A., Batie, S., 1978. Economic value of natural coastal wetlands: a critique. Coastal Zone Management Journal 4, 231-247.
Shaw, W. D., Jakus, P., 1996. Travel cost models of the demand for rock climbing. Agricultural and Resource Economics Review 25 (2), 133-142.
SIKA, Tid och kvalitet i persontrafik. SIKA rapport 2002:8. SIKA, Stockholm.
Smith, V. K., 2004. Fifty years of contingent valuation, pp. 1-60 in: Tietenberg, T., Folmer, H. (eds.), The International Handbook of Environmental and Resource Economics 2004/2005: A Survey of Current Issues. Edward Elgar Publishing, Cheltenham, UK.
Smith, V. K., in preparation. Judging quality. In: Kanninen, B. (ed.), Valuing Environmental Amenities using Choice Experiments: A Common Sense Guide to Theory and Practice, forthcoming.
Söderqvist, T., 1996. Ekonomisk värdering av miljön: metoder och svenska erfarenheter, pp. 5-73 in Expertrapporter från Skatteväxlingskommittén, SOU 1996:117. Finansdepartementet, Stockholm.
Söderqvist, T., 1998. Why give up money for the Baltic Sea? Motives for people's willingness (or reluctance) to pay. Environmental and Resource Economics 12, 249-254.
Söderqvist, T., Hammer, M., Gren, I-M., 2004. Samverkan för människa och natur – en introduktion till ekologisk ekonomi. Studentlitteratur, Lund.
Sundberg, S., 2004. Replacement costs as economic values of environmental change: a review and an application to Swedish sea trout habitats. Beijer Discussion Paper Series No. 184, Beijerinstitutet för ekologisk ekonomi, Kungl. Vetenskapsakademien, Stockholm.
Sundberg, S., Söderqvist, T., 2004a. ValueBaseSWE: a valuation study database for environmental change in Sweden. Available at www.beijer.kva.se/valuebase.htm. Beijer International Institute of Ecological Economics, The Royal Swedish Academy of Sciences, Stockholm.
Sundberg, S., Söderqvist, T., 2004b. The economic value of environmental change in Sweden: a survey of studies. Rapport 5360, Naturvårdsverket, Stockholm.
Svenska Statistikersamfundet, 2005. Standard för bortfallsberäkning. Sektionen för survey-statistik (www.statistikersamfundet.se/survey), Svenska Statistikersamfundet.
USEPA, 2000. Guidelines for Preparing Economic Analyses. EPA 240-R-00-003. United States Environmental Protection Agency, Washington, DC.
Wärneryd, B. (red.), 1990. Att fråga – om frågekonstruktion vid intervjuundersökningar och postenkäter, femte upplagan. Statistiska Centralbyrån, Stockholm.
Willig, R. D., 1976. Consumer's surplus without apology. American Economic Review 66, 589-597.

APPENDIX A

Valuation methods[2]

Valuing environmental changes economically is about analysing the trade-offs individuals are prepared to make between the environment and other resources. Economic theory suggests that such trade-offs reveal the influence that environmental changes have on human wellbeing. In other words, economists measure the influence of an environmental change on wellbeing as the resources individuals would be willing to give up in order to have the change (or prevent the change). Another word for this willingness to give up resources is *willingness to pay* (WTP). In some situations it is more relevant to study another kind of trade-off, namely what people require as compensation if the environmental change takes place (or is prevented), i.e. their *willingness to accept compensation* (WTA); see, e.g. Freeman (2003). The change in producer surplus is the corresponding measure for changes in firms' "wellbeing".

Since economic values are about trade-offs that individuals are willing to make, the economic value will depend on the individuals' preferences, i.e. their more or less fixed opinions on how important (or unimportant) different issues are to them. The underlying factors playing a role for the creation of these opinions tend to be studied by psychologists rather than economists, but it may be noted that it has been asserted, inter alia, that individuals take on different roles, e.g. as citizens or consumers, and that this may influence their behaviour, see, e.g. Sagoff (1988). The focus on individuals' preferences in economics is an effect of a purely anthropocentric ethical point of departure, and also of the importance of the principle of consumer sovereignty, i.e. that every individual is the sole person who can judge what is good or bad for her. A discussion of such points of departure is beyond the scope of this report, but it should be noted that the view that economic values are determined by individuals' preferences implies that the results from valuation studies are not more informed than the individuals themselves are (Daily et al. 2000). This fact has probably played an important role for the discussion among, for example, natural scientists and economists about the reasonableness of economic valuation of environmental change.

Sometimes environmental changes result in effects regarding goods and services that are subject to free trade and pricing on markets. An example might be an afforestation project resulting in benefits in terms of an increased supply of the market good of timber. If there is enough information about demand and supply on the timber market, the economic value of the project can be estimated as the resulting changes in consumer surplus and producer surplus.

A more difficult case occurs when the effects influence the supply of goods and services provided by nature, or "ecosystem services" as they are often called

[2] The text in this appendix follows Söderqvist et al. (2004).

today. We use this term below in the widest possible sense. The difficulty arises due to the fact that many ecosystem services are not subject to trade on any market. To conduct the valuation, special valuation methods have been developed. These valuation methods may be divided into three main groups:

1. Revealed preferences methods (RP).
2. Stated preferences methods (SP).
3. Other valuation methods (less firmly rooted in economic theory).

In what follows we briefly go through the valuation methods found in each one of these main groups.

A.1 Revealed preferences methods

The methods in this group use linkages that exist between ecosystem services and one or more market goods. The four most important valuation methods within this group are:

- The production function method (PF).
- The travel cost method (TCM).
- The property value method (often called the hedonic price method) (HP).
- The defensive expenditure method (DE).

The production function method may be applied when ecosystem services are used in the production of some market good. Ecosystem services are often such an input. For example, the production in agriculture and forestry depends on soil fertility, which in turn is maintained by the work carried out by various organisms in the soil. The harvest is then subject to trade on a market, which can be described by demand and supply curves. Another example is how cod fishing in the Baltic Sea is dependent of marine water quality. The fishery industry is able to catch (”produce”) fish thanks to, inter alia, labour, tools and ships as well as a number of marine environmental factors. The cod caught is a market good for which demand and supply curves may be estimated. If it is possible to establish how demand and supply is influenced by a change in the supply of the ecosystem service, it is also possible to value this change economically. The production function method is an important valuation method, but its application is often limited by insufficient knowledge of how nature works as a production factor.

The travel cost method provides an opportunity to value the recreational opportunities that nature offers. The willingness to pay for visiting a recreational area may be estimated if there are enough data on how much money and time people spend in order to travel to the area. The original application of the travel cost method was to study the recreational access value of areas such as nature reserves in the US. The idea of a more modern version of the method is to

analyse how different properties of a recreational area affect the demand for recreation. It may, for example, concern the assumption that the water quality by a beach plays a role for how many people will visit the area. If knowledge is available of how water quality is manifested, and if the effect of water quality on recreational demand can be isolated from that of all other factors influencing the demand (travel cost, income, services on site, etc.), there exist possibilities to derive the willingness to pay for improved water quality.

The property value method departs from the idea that the supply of the ecosystem service may play a role for house prices. A summer house situated by a beach with poor water quality may have a lower market price than a summer house situated by a beach with clean water, even if the houses and the surroundings in all other respects are identical. If data exist on house prices, characteristics and surroundings, water quality included, an indirect market price on water quality may be estimated and in some cases even the willingness to pay for an improved water quality.

The defensive expenditure method uses data on individuals' market behaviour when they have the purpose of compensating themselves for deteriorated environmental quality or reduced supply of some ecosystem service. One example is when people install some equipment to protect themselves from a deterioration of the environment, for example, a coal filter cleaning the drinking water coming from contaminated groundwater. From a drinking water perspective such a filter works as a substitute for clean groundwater if the filter preserves the quality of the drinking water. For a small change in the supply of an ecosystem service, such defensive expenditures may give information on the willingness to pay for the change.

A.2 Stated preferences methods

Sometimes there is no linkage between the ecosystem service one wishes to value and some market good, or the linkage is weak or poorly explored. With the help of stated preferences methods, this problem can be solved by estimating the WTP for the ecosystem service directly. One might say that stated preferences methods are all about creating hypothetical market situations. This way of gaining information about the economic value of the environment has been increasingly applied during the last decades. There are a number of stated preferences methods, but we simplify by saying that the two main methods are the following:

- The contingent valuation method (CVM).
- Choice experiments (CE).

The contingent valuation method uses interviews or mail surveys that describe a scenario where a change in the supply of an ecosystem service is explained and

illustrated for a (usually) randomly selected sample of individuals. Next, questions are posed about the individuals' willingness to pay for a realisation of the change. This is a debated method, not least among economists (cf. appendix B1). The requirements are substantial regarding the design of text and pictures, as well as other things conveying the considered change of the ecosystem service. But most controversial for economists is probably the fact that the method does not use data on individuals' actual behaviour on some market. A main question is whether individuals would actually pay the WTP expressed by them at the hypothetical market if the scenario becomes a reality. However, the hypothetical nature of the CVM makes it possible to reveal the valuations of people who do not use the ecosystem service being valued, so-called *non-use values*. For example, only values held by visitors are taken into account if an improved environmental quality in a recreational area is valued with the travel cost method. However, it is not unlikely that also non-visitors care for the environmental quality in the area. A CVM study can be used for capturing these non-visitors' valuations.

Choice experiments resemble the CVM, but is based on how the selected individuals make repeated choices among at least two alternatives. The alternatives differ with respect to levels of attributes characterising the environment and the payment requirements for the respondent. A willingness to pay for the environmental attributes can be derived from the choices made by the respondents.

A.3 Other valuation methods

A joint feature of the methods mentioned so far is that they can all be justified by economic theory. There are however other methods that are also used for valuing environmental changes, but they are not as firmly rooted in economic theory. This may make the interpretation of the results difficult. We will briefly describe three of these methods:

- The replacement cost method (RCM).
- The human capital method (HCM).
- The costs of realising political decisions ("political WTP", pWTP).

The replacement cost method resembles the defensive expenditure method in the sense that it is applied to cases where a market good might replace an ecosystem service. But when the replacement cost method is used, the costs for socially co-ordinated actual or hypothetical projects are studied rather than individuals' actual trade-offs at a market. One example may be a study of the costs for building flooding protection along rivers in order to at least partly compensate for the increasing variation in water flows that logging and ditching of wetlands may result in. Another example is the construction of sewage treatment plants to compensate for the lost water cleaning capacity when wetlands are ditched. Replacement costs refer to the costs of replacing the loss of an ecosystem service.

These costs can be interpreted as the economic value of the ecosystem service, given that the following conditions hold: (i) the man-made replacement system provides services of equally high quantity and quality as the ecosystem service, (ii) the man-made replacement system is the cost-effective way of replacing the ecosystem service, and (iii) citizens would in fact be willing to pay the costs for the replacement system if the ecosystem service is no longer available. Note that if the second and third conditions are fulfilled, but not the first condition, the replacement cost method tends to lead to underestimations of the economic value. On the other hand, the method would lead to overestimations of the economic value if the first and third conditions are fulfilled, but not the second condition.

The human capital method is based on, inter alia, the idea that a person's value is what she produces and that wages give information about productivity. The method provides a basis for using data on production losses in order to value illness. To this is usually added calculations of costs of medical treatment. Such data are interesting and may in some cases be motivated by the fact that they provide information on the lower boundary of economic damage. However, the method has to be used cautiously because it may give results that are not defensible, for example, that retired people have no value.

The cost of realising political decisions provides some valuation possibilities by using cost data. It is doubtful whether such decisions reveal "society's willingness to pay" for a changed supply of ecosystem services since the decisions do not necessary reflect the citizens' opinions. However, this does not imply that there are no cases where citizens' opinions are relatively strongly reflected. The decision was perhaps *preceded* by an intensive discussion in which the opinions of many groups were expressed and also converged. A valuation through the cost of realising political decisions has some similarities with the replacement cost method, which suggests that the three conditions mentioned above for the RCM are again applicable.

A.4 The use of valuation methods in Sweden

All valuation methods described above have been used in a Swedish context. About 100 Swedish valuation studies are summarised in detail in the database ValueBase^{SWE} (see appendix B1). 68 per cent of these studies used some SP method, 21 per cent some RP method and 11 per cent used some other valuation method (Sundberg and Söderqvist 2004b). Table A1 shows that the contingent valuation method was the most common SP method. The travel cost method was the most applied RP method.

Table A.1. *The use of valuation methods in Sweden.*

Method		Per cent within main group of methods	Per cent overall
1.	*Revealed preferences methods*		*21*
	The production function method	11	
	The travel cost method	45	
	The property value method	33	
	The defensive expenditure method	11	
2.	*Stated preferences methods*		*68*
	The contingent valuation method	82	
	Other SP methods, e.g. choice experiments	18	
3.	*Other valuation methods*		*11*

Source: Sundberg and Söderqvist (2004b).

APPENDIX B1

Quality of valuation studies – earlier results

This appendix presents some selected results from the literature on the quality of valuation studies. Literature database and internet searches, inquires at the environmental economics mailing list RESECON and contacts with key scholars such as Reed Johnson, Michael Hanemann, Ståle Navrud, Kerry Smith, and the Reference Group of this project indicate that there is at present few comprehensive compilations of the kind of operational quality criteria that this report is aiming at. However, some important work seems to be ongoing, for example by Smith (in preparation) about evaluation of choice experiments. The existing literature still partly indicates how quality criteria might be designed. Such indications from three different sources are presented below:

1. Scientific literature
2. Guidelines prepared by authorities
3. Databases for valuation results

B.1.1 Quality criteria in scientific literature

When quality of valuation studies are discussed, many readers probably first relate to the periodically intensive discussion about the validity of the contingent valuation method. Economists have generally tended to be sceptical to the use of mail questionnaires or interviews for collecting economic data, in particular data about economic values. Several scholars have argued that individuals will behave strategically in such situations and thus not reveal preferences truthfully. For example, the following judgement by Paul Samuelson (1954) has probably stimulated this scepticism:

> "It is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective activity than he really has. – One could imagine every person in the community being indoctrinated to behave like a 'parametric decentralized bureaucrat' who reveals his preferences by signalling in response to price parameters or Lagrangean multipliers, to questionnaires, or to other devices. Alas, by departing from his indoctrinated rules, any one person can hope to snatch some selfish benefit in a way not possible under the self-policing competitive pricing of private goods." (cited in Mitchell and Carson 1989, pp. 127–128)

Later experimental economics research has qualitatively confirmed Samuelson's judgement in the sense that individuals tend to behave strategically when they have incentives to do so, but the magnitude of this behaviour is far from always of the size predicted by economic theory (Davis and Holt 1993). The presence of strategic behaviour should thus not à priori preclude the use of mail questionnaires and interviews for collecting information about economic values. The risk for strategic behaviour might instead be viewed as one of several problems that the CVM and other SP methods have to consider in order to minimise their influence on individuals' behaviour when responding to a mail questionnaire or an interview.

The contingent valuation method entered in the research community in earnest in the 1970's, when CVM applications began to be published in environmental economics journals, for example, Randall et al. (1974) and Brookshire et al. (1976). However, the method would develop from a research tool used for scientific experiments to a method used in real policy contexts – "from esoteric toy to multibillion dollar assessment tool" (Kriström 1996, p. 2). A book by Robert C. Mitchell and Richard T. Carson (1989) became instrumental for the penetration of the method. The book presented and discussed the method thoroughly, and it also contributed to a standardisation of the method by nailing the term of "the contingent valuation method". One of the first detailed suggestions of how a CVM study might be evaluated is also found in the book, see table B.2.1 in appendix B2.

An event that contributed strongly to give attention to the method happened in the same year that the book by Mitchell and Carson was published. On 24 March, 1989, the tanker *Exxon Valdez* run aground off the coast of Alaska and birds and other organisms died because of the oil that leaked out. The accident resulted in a legal action about damages between the State of Alaska and Exxon Corporation. In order to obtain a basis for the case, the State of Alaska funded a CVM study estimating the extent of the economic damage caused by the accident (Carson et al. 1992). Non-use values turned out to play an important role in this study, which was carried out by a number of distinguished environmental economists. On the other hand, Exxon funded CVM research that not surprisingly happened to illustrate weaknesses associated with CVM (Hausman 1993). Exxon also involved top-level economists, but some of them were rather new in the environmental economics arena. The big monetary amounts at stake stimulated an inflamed discussion, and the US National Oceanic and Atmospheric Administration initiated an evaluation of the validity of CVM as a valuation method by an independent committee (Arrow et al. 1993). A conclusion of this blue ribbon panel was as follows:

> "CVM studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values. To be acceptable for this purpose, such studies should follow the guidelines described." (p. 4610).

The guidelines suggested by the NOAA panel are found in table B.2.2 in appendix B2. The conclusion was thus encouraging for the CVM in the sense that the panel in principle approved the CVM as a valid valuation method. Note also that the panel accepted the view that non-use values are an economic value. On the other hand, the guidelines imply that CVM studies have to be extremely carefully designed and implemented. This makes them expensive. Fulfilling the requirements in the guidelines in all respects would imply survey costs high enough to put almost all potential CVM applications on the shelf. According to the panel, CVM studies should, for example, not make use of mail questionnaires, but face-to-face interviews are to be preferred. Smith (2004) notes that while the panel did not explicitly ban CVM studies, the guidelines imply that the panel "priced the practice out of the market" (p. 16). Moreover, Smith argues that only one CVM study has been carried out with the purpose of following the guidelines in all respects. This illustrates that the panel's recommendations have to be interpreted in the light of the State of Alaska vs. Exxon case. The recommendations also reflect conditions that might be valid in the US, but not necessarily in other countries. For example, mail questionnaires carried out in Sweden are considerably more likely to result in satisfactory response rates and representativity of a population than the case is in the US. It is therefore not reasonable to view the panel's recommendations in table B.2.2 as a generally valid law for how CVM studies have to be designed. However, the recommendations indicate what CVM characteristics might influence the validity of the CVM as a valuation method.

The discussion about the CVM has resulted in extensive research about the potential weaknesses of CVM. According to Smith (2004), the CVM has in this way caused the most thorough investigation of individuals' preferences ever undertaken in economics. The big number of tests of various CVM characteristics has contributed to a greater understanding for in what circumstances the method probably works well and when it works poorly (Carson et al. 2001). While the CVM discussion is more balanced today, some economists still argue that SP methods do not give useful information about economic values. The main reasons are the hypothetical nature of SP methods and these methods' potential to capture non-use values, which some economists do not regard as valid economic values (Smith 2004, USEPA 2000). However, we believe that economists arguing that SP methods should be dismissed entirely are in the minority. For example, according to Haab and McConnell (2002),

> "[…in recent years,] stated preferences methods have become more accepted. The debate about valuation by stated preferences is over, with the possible exception of its use in eliciting existence values. Contingent valuation has proved to be no less reliable than behavioural methods in a variety of tests." (p. 3)

Table B.1.1 provides an example from more recent literature about evaluation of the quality of CVM studies (Carson 2000). Some other examples are found in tables B.2.3–B.2.5 in appendix B2. Several recommendations made by Carson (2000) are similar to those suggested already by Mitchell and Carson (1989). However, there are also some differences, and it can be noted that Carson (2000) is more general in his recommendation to analyse how sensitive CVM results are for the choice of econometric model. This probably reflects that in the 1990's, open-ended WTP questions and the use of payment cards were to a large extent replaced by preference elicitation by asking respondents to accept or not accept to pay a given amount of money. Whether such a discrete choice elicitation format actually performs better is subject to discussion. The fact that discrete choice questions resemble a "normal" market situation is one of the probable advantages. However, since they only result in a "yes" or "no" answer, they give little information about the actual WTP. This might result in wide confidence intervals for value estimates. For example, estimates of mean WTP have been shown in some cases to be very sensitive with respect to assumptions on the probability distribution for the mean WTP (Haab and McConnell 2002).

A recommendation that is constantly repeated is that a questionnaire to be used in an SP study must be very carefully tested and designed, cf. Carson (2000) in table B.1.1 and Bishop (2003) in table B.2.3. No econometric analysis, its degree of refinement notwithstanding, is likely to provide a remedy for an unclear valuation scenario, a high non-response rate or many protest answers.

In summary, an important conclusion from the CVM discussion seems to be that many of the problems associated with the method can be solved or avoided if the study is carefully designed and implemented. How this can be accomplished is indicated by the advice in table B.1.1 and tables B.2.3–B.2.5. It thus seems as if the validity of CVM as a valuation method to some extent is a matter of the availability of funds, since it might be costly to carry out, for example, detailed pretests. Carson et al. (2001) note that it is therefore a need for research on how survey costs can be cut without reducing study quality too much.

Table B.1.1. *Questions to consider when evaluating a CVM study.*

A. The survey instrument

1. Is there face validity? For example, is the good and scenario under which it would be provided described clearly and accurately? Is the trade-off that the respondent is asked to make a plausible one? Is the respondent provided with enough information to make an informed decision without being overwhelmed with it?
2. Does the survey instrument include the following: An introductory section that helps set the general context for the decision to be made? A detailed description of the good to be offered to the respondent? The institutional setting in which the good will be provided? The manner in which the good will be paid for? A method by which the survey elicits the respondent's preferences with respect to the good? Debriefing questions about why respondents answered certain questions the way that they did? A set of questions regarding respondent characteristics including attitudes and demographic information?

B. Developing the survey instrument

1. Did the development work include focus groups and in-depth interviews for determining the plausibility and understandability of the good and the scenario being presented?
2. Were pretests and pilot studies carried out to assess how well the survey works as a whole?

C. Sampling and population

1. Was the particular population sampled relevant for evaluating the benefits and/or costs of the proposed project?
2. Was the sample size at least 300-2000? Such a sample size is generally required to achieve reasonable reliability from a sampling perspective because survey data are typically highly variable when trying to measure a continuous variable.
3. Did all members of the relevant population have a positive and known probability of being included in the sample?
4. Was an appropriate set of weights used if inclusion probabilities are not equal?

D. Survey administration and response rate

1. Were in-person interviews used? If so, were professional interviewers used? In-person interviews generally facilitate understandability because visual materials such as maps and pictures can be used.
2. If a mail survey was used, was the presence of potential sample selection bias investigated? Was any technique used for correcting for sample selection bias?
3. Was there a high response rate (60-80%)? If so, potential problems with extrapolating to the population of interest are minimized.
4. How were non-respondents treated? Were any procedure used for minimizing non-response bias?

E. The scenario

1. Was the selected unit of observation appropriate? Households are likely to be the appropriate unit if a payment vehicle such as higher taxes or utility bills are used. Individuals are likely to be appropriate if, for example, entrance fees were the payment vehicle.
2. Was the payment described as a lump sum or a continuing payment? A one-time payment generally produces more conservative estimates since it does not offer the opportunity to spread payments over time. A one-time payment is appropriate in cases where providing the good represents a one-time event, but not in cases for which ongoing easily visible actions must be taken.
3. Was the respondent asked for information about WTP or WTA? WTA questions are usually much harder to successfully implement, but they often represent the correct property rights perspective.

F. Data analysis

1. Were inevitable judgmental decisions about handling and analysing the data adequately reported?
2. Was a valuation function estimated? If so, did it have a reasonable explanatory power and coefficients with the expected signs?
3. How were outliers and protest answers treated?
4. How sensitive were results for the choice of econometric model?
5. Was the distribution of economic value on a per-capita basis reasonable? For most environmental goods, WTP distributions based on the general population (not specific populations such as hunters) will be quite asymmetric with mean WTP larger than median WTP, in part because the income distribution is asymmetric and in part because there is often a sizable part of the population that is fairly indifferent to the environmental good and a smaller group that care a great deal about its provision.
6. Was the correct WTP measure reported with respect to how it is intended to be used? Mean WTP is the traditional measure used in cost-benefit analysis, while median WTP is a standard public choice criterion.

Source: Carson (2000).

In recent years, CE studies have gradually become more common as a valuation method. CE and CVM studies share many characteristics, but their differences have initiated a discussion about for which applications CE studies are more suitable than CVM studies. It is often suggested that CE studies are more adequate to use when it is reasonable to describe an environmental resource by using a number of attributes and when one is interested in the value of marginal changes in these attributes. The analysis of CE data usually involves rather restrictive assumptions for the underlying utility model, and respondents face a task that is at least not less difficult than in CVM studies when it comes to understanding the valuation scenario and answering preference elicitation questions. See, e.g. Alpízar et al. (2003) and Bateman et al. (2002).

There has been an enormous attention paid to the contingent valuation method in comparison to other valuation methods. This is understandable because of the hypothetical nature of the valuation questions posed in a CVM study, but the difference in attention is to some extent unjust. The reason is that also the reliability of RP methods depends on good data quality. RP methods often involve data collection by surveys, which means that also the application of RP methods can be improved by taking into account the advice available for CVM studies. Bishop (2003) notes that:

> "…the confidence we economists have in the logic of revealed preference has lulled us into complacency about validity issues. Starting with revealed preferences data is no guarantee of validity." (p. 588)

> "…revealed preferences studies need to give more attention to validity. They are no different than stated preferences studies in this regard." (p. 560).

Table B.1.2 presents Bishop's suggestions for further research in order to increase the reliability of RP methods. For the particular case of the travel cost method, Phaneuf and Smith (2004) emphasise that it would gain from more research on the opportunity costs of travel time and on-site time, intertemporal restrictions (and opportunities) for individuals' choices, definitions and measurements of environmental quality variables and the type of recreation produced and consumed, and the treatment of multipurpose trips.

Table B.1.2. *Some important ways to test and increase the reliability of RP methods, according to Bishop (2003).*

1. Content validity: while a tremendous amount of effort has gone into improving econometric methods for RP studies, work on data quality is stunted. Survey methodology for RP studies is a neglected area.
2. Convergent validity: in cases when RP studies involve survey work, stated preference questions should be included for comparison with RP results.
3. Criterion validity: try to design research so that environmental quality becomes a real market good, and compare those market results to results of RP applications. It might be possible to raise enough money to gather very high quality data on recreationists or home buyers and sellers in order to test the methods we normally use in travel cost and hedonic studies.

B.1.2 Guidelines prepared by authorities

The increased significance of valuation studies in policy contexts has made it important for authorities to prepare guidelines for how valuation studies should be applied in order to have a satisfactory quality. Such guidelines indicate quality criteria for which the user dimension probably is relatively important. We use the US Environmental Protection Agency's *Guidelines for Preparing Economic Analyses* from September 2000 as a main example of guidelines. A scientific reference group consisting of distinguished American environmental economists made an evaluation and concluded that the guidelines reflect methods and techniques being generally accepted among environmental economists. The tables below present questions that according to USEPA should be used for evaluating valuation studies. The questions are specific for different valuation methods and concern the production function method (table B.1.3), the travel cost method (table B.1.4), the property value method (table B.1.5), the defensive expenditure method (table B.1.6), the contingent valuation method (table B.1.7) and the human capital method (table B.1.8). Any guidelines for evaluating CE studies were not included in USEPA (2000) because CE were at that time a rather new method for environmental valuation. Tables B.2.6 and B.2.7 in appendix B2 give two additional examples of guidelines for the use of valuation methods prepared by authorities in the UK and Australia.

Table B.1.3. *Important questions to consider when evaluating studies using the production function method.*

A. Data requirements and implications
1. Was information available on the effect of the environmental resource on production costs?
2. Was information available on supply conditions for output?
3. Was information available on demand curve for final good?
4. Was information available on factor supplies?
B. The model for estimation
1. What modelling approach and structure of the model were selected? Data availability plays a large role in this selection. Production function, cost function and simulation and optimization models are all options for understanding the market response to environmental improvements.

Source: USEPA (2000, p. 73).

Table B.1.4. *Important questions to consider when evaluating studies using the travel cost method.*

A. Definition of a site
1. What was the compromise in the recreation demand study in defining sites, balancing data needs and availability, costs, and time? Ideally, one could estimate a recreation demand model in which sites are defined as specific points, such as launch ramps, campsites, etc., but the data requirements of detailed models are large. Similarly, for a given site, the range of alternative sites may vary by individual?
B. Opportunity cost of time
1. How was the value of recreation time defined and estimated?
2. Was it assumed that travel time detract from the overall satisfaction of a recreation trip? If so, was this assumption reasonable?
3. How was on-site time treated?
C. Multiple site or multipurpose trips
1. Was it assumed that the particular recreation activity being studied is the sole purpose for a given trip? If so, was this assumption reasonable? If not, how was the issue of multipurpose trips approached?
2. Was it assumed that the particular recreation activity involved the visit to one single site? If so, was this assumption reasonable? If not, how was the issue of multiple site approached?

Source: USEPA (2000, p. 74).

Table B.1.5. *Important questions to consider when evaluating studies using the property value method.*

A. Data requirements and implications

1. Were market transaction prices on individual parcels or housing units used as data? Such data are preferred to aggregated data such as census tract information on average housing units because aggregation problems can be avoided.
2. What attributes were there data on? Such data may include housing characteristics, sale dates, neighbourhood amenities such as schools and parks, neighbourhood demographic characteristics such as income, age, and race, and environmental quality.

B. Errors in variables

1. Were there problems due to errors in measuring prices (aggregated data)?
2. Were there problems due to errors in measuring product characteristics (e.g. those related to the neighbourhood and the environment)?
3. Were there any omitted variable bias problems? Such problems may occur if relevant data are not available.

C. The measurement of environmental attributes

1. Did the study use information available from the scientific community for measuring environmental attributes? If so, were there differences between how environmental attributes are measured by scientists and how they are perceived by individuals? If this difference is large, the hedonic price function will not accurately represent the values of these attributes. Individual perceptions of environmental attributes are central to this type of analysis.
2. What was the timing of the effect from the environmental change? Some effects from environmental change vary over time. Others may be understood differently over time depending on available information (e.g. hazardous waste sites).

D. The model for estimation

1. Was there an analysis of the implications in terms of benefits assessment of the choice of the functional form of the estimated hedonic price function? Economic theory offers limited guidance on the functional form of a hedonic price function. However, the choice of functional form has implications for benefits assessment.
2. How was the extent of the housing market defined? It is important to note that if the market is defined to be too big, the resulting coefficients of the hedonic price function may be biased. Conversely, if the market is defined too narrowly, the coefficients of the hedonic price function are less efficient.

E. Evaluation of the results

1. Was the empirical work reviewed? Such a review would include assessments of (i) the quality of the data collected, (ii) the framing of the policy problem, (iii) the measurement of environmental attributes, and (iv) the statistical regression analysis.
2. Was there a comparison to earlier results in the literature? Comparing data, modelling assumptions, and results across studies is a useful exercise. While variation is expected across studies, especially those completed on different areas, some factors such as the signs of particular coefficients may be consistently reported.

Source: USEPA (2000, pp. 78-79).

Table B.1.6. *Important questions to consider when evaluating studies using the defensive expenditure method.*

A. Data requirements and implications
1. Were there enough data for estimating WTP? The data requirements are quite burdensome and include information detailing the severity, frequency, and duration of symptoms, exposure to environmental contaminants, actions taken to avert or mitigate damages, the costs of those behaviours and activities, and other variables that affect health outcomes (e.g. age, health status, chronic conditions). Often, data availability will limit the analysis to an examination of observed defensive expenditures. These results can be cautiously interpreted as a lower bound on WTP. Note that costs associated with pain and suffering will not be included in the estimate.
2. Were there big differences between perceived benefits from defensive behaviour and objective estimates of, for example, risk changes? If so, the analysis will produce biased WTP estimates for a given change in objective risk. Surveys may be necessary in order to determine the benefits individuals perceive they are receiving when engaging in defensive activities. These perceived benefits can then be used as the object of the valuation estimates.
B. Accounting for other benefits and disutilities
1. Did the defensive behaviour provide other benefits than mitigation against environmental damages? If so, were these benefits disentangled? In order to accurately produce estimates of WTP for a risk change, for example, averting behaviour studies must isolate the value for the effect of interest from the value of the other benefits conferred by the defensive activity.
2. Did the defensive behaviour have any negative effects on utility? For example, wearing helmets when riding bicycles may be uncomfortable.
C. Modelling assumptions
1. Is the modelling based on an assumption that the economy and the environment are additively separable? This assumption may lead to unambiguous results, but it may be plausible only in particular circumstances and should therefore be justified whenever invoked. There is a need to review and assess the implications of this and other assumptions for the valuation estimates.

Source: USEPA (2000, pp. 80-81).

Table B.1.7. *Important questions to consider when evaluating studies using the contingent valuation method.*

A. Content validity
1. Was the commodity being valued clearly and concisely defined? A detailed explanation of the salient features of the environmental change being valued ("the commodity") begins with a careful exposition of the conditions in the baseline case and how these would be expected to change over time if no action were taken.
2. Was the policy change adequately described? This description should include an illustration of how and when the policy action would affect aspects of the environment that people might care about.
3. What was the rationale for the choice of payment mechanism? The way the payment will be made (e.g. through taxes, user fees, etc.) may have large implications for the outcome.
4. Was the scenario comprehended and accepted by the respondents? Respondent attitudes about the provider and the implied property rights of the survey scenario can be used to evaluate the appropriateness of these features of the commodity description. Questions that probe for respondent comprehension and acceptance of the scenario can offer important indications about the potential for the study to be reliable.
B. Construct validity
1. Were there tests of internal validity? If so, what were the results? Internal validity is supported when variables that are expected by theory to be important determinants of preferences actually are statistically significant with the correct sign. For example, with normal goods, price is expected to have a negative effect on demand for a good, while household income is expected to have a positive effect, all else equal.
2. Were respondents familiar with the good or its context? One would expect that someone who fishes would know more about, and be willing to pay more for, a commodity that improves conditions for fishing than someone who never engages in outdoor recreation.
3. Was there sensitivity to scope? Scope tests, where the amount of the commodity is varied randomly over different sub-samples of survey respondents, can increase confidence in the results where the findings are consistent with theoretical expectations.
C. Criterion validity
1. Could the CVM study estimates be compared to indicators of true value? Given the lack of actual market prices, it is often impossible to conduct criterion validity tests. However, the quality of a CVM study can also be gauged by comparing valuation estimates obtained using CVM with those obtained using other techniques.

Source: USEPA (2000, p. 85).

Table B.1.8. *Important questions to consider when evaluating studies using the human capital method.*

A. Theoretical considerations
1. How are the results interpreted in terms of WTP? The cost of illness is not a measure of WTP to avoid an illness but rather a measure of the ex post costs of an illness. The approach relies on the two major assumptions that (i) direct costs of morbidity reflect the economic value of goods and services used to treat illness, and (ii) a person's earnings reflect the economic value of lost production. The approach simply measures ex post costs and does not attempt to measure the loss in utility due to pain and suffering or the costs of any averting behaviours that individuals have taken to avoid the illness altogether. Also, ex post measures cannot capture any value associated with risk attitudes. However, the cost-of-illness estimate may be considered a lower bound estimate of WTP.
B. Technological change
1. Have there been big changes in medical treatment technologies and methods? Such changes could push the true cost estimate for a given illness either higher or lower. When using previous cost-of-illness studies, the analyst should be sure to research whether and how the generally accepted treatment has changed from the time of the study.
C. Measuring the value of lost productivity
1. How was the loss in work time estimated? Simply valuing the actual lost work time due to an illness may not capture the full loss of an individual's productivity in the case of a long-term chronic illness. Chronic illness may force an individual to work less than a full-time schedule, take a job at a lower pay rate, or drop out of the labour force altogether.
2. How was the cost of time defined? Even if the direct medical costs are estimated using individual actual cost data, it is highly unlikely that the individual data will include wages. Therefore, the wage rate chosen should reflect the demographic distribution of the illness under study. Furthermore, the value of lost time should include the productivity of those persons not involved in paid jobs, and preferably also the value of lost leisure time.

Source: USEPA (2000, pp. 82-83).

B.1.3 Quality criteria in databases

The increasing number of valuation studies has called for compilations of valuation results, in particular user-friendly databases. A crucial question when setting up such a database is whether and how some kind of quality labelling should be made for the studies that are included in the database. One type of labelling could be to only include studies that fulfil the requirements for what is judged to be a minimum acceptable quality. We report below about the structure of four databases and to what extent these include some type of quality evaluation. The databases are the Environmental Valuation Reference Inventory (EVRI), the Envalue Environmental Valuation Database (ENVALUE), Valuation Study Database for Environmental Change in Sweden (ValueBase^{SWE}), and the New Zealand Non-Market Valuation Database.

B.1.3.1 THE ENVIRONMENTAL VALUATION REFERENCE INVENTORY (EVRI)

The most extensive database of valuation studies is the Environmental Valuation Reference Inventory (EVRI), see www.evri.ca. In September 2004, EVRI included information on about 1400 valuation studies, about 800 of which from the US or Canada (and about 20 from Sweden). Slightly more than 50 per cent of the studies have made use of some SP method. The database is accessible free of charge for citizens in Canada, France, the UK and the US, which are the countries that have undertaken to contribute to further development of the database. Users from other countries pay CAD 900 for a 12-month subscription or CAD 200 for a 1-month subscription.

EVRI contains detailed information about each included valuation study. The information is structured into more than 30 data fields, divided into the following six categories:

1. *Study reference:* basic bibliographic information.
2. *Study area and population characteristics:* information about the location of the study along with population and site data.
3. *Environmental focus of the study:* fields that describe the environmental asset being valued, the stressors on the environment, and the specific purpose of the study.
4. *Study methods:* technical information on the actual study, along with the specific techniques that were used to arrive at the results.
5. *Estimated values:* the monetary values that are presented in the study as well as the specific units of measure.
6. *Alternative language summary:* an abstract of the study available in English, French and Spanish.

An explicit purpose of this detailed information is to facilitate users' search for studies that might be suitable for benefit transfers (also called value transfers), i.e. generalising existing valuation results to new situations. An identification of

potentially suitable studies for benefit transfer is made on the basis of similarity between the existing study (*study site*) and the new situation that is to be subject to valuation (*policy site*) for the following areas:

- Geographic location
- Population
- Environment
- Timeliness of data
- Economic measure
- Estimated values
- Abstract
- Complete study

The information in EVRI is without doubt helpful for sorting out valuation studies that might be suitable for benefit transfer, but it is up to the user to judge the quality of the existing study. The information given in EVRI does thus not include any explicit quality judgement. Quality assessments have been discussed among those responsible for EVRI ever since the planning for the database began more than ten years ago. The difficulty to make an objective quality assessment, and the quite sensitive issue of assessing studies whose authors EVRI relies on for its future development, have refrained EVRI from quality assessments (McComb 2004). However, the selection of studies to be included in EVRI involves a check on quality. Reports and working papers being obviously poorly written are sorted out, and papers published in scientific journals are assumed to have a satisfactory quality. However, this procedure is not entirely reasonable since journal articles might not be very policy relevant. For the sake of usefulness, it is likely to be important to also include "the grey literature" with reports from, e.g. authorities and consultancies (McComb 2004).

B.1.3.2 THE ENVALUE ENVIRONMENTAL VALUATION DATABASE (ENVALUE)

ENVALUE (www.epa.nsw.gov.au/envalue) is an Australian database developed by the New South Wales Environmental Protection Agency. It is free of charge to use, but it is in several respects less detailed than EVRI and it includes primarily information about Australian valuation studies (about 130 studies, November 2004). A number of studies from other countries are also found in the database, including 200 studies from the US and three Swedish studies.

The following information is found in ENVALUE about the valuation studies:

- Study information: basic bibliographic information.
- Environmental medium: e.g. natural areas with wilderness as a sub-medium and forests as a sub-submedium.
- Country and location (state and site).
- Year of data.

- Attribute measured.
- Units of measurement.
- Valuation method.
- Key results including estimated values.
- Dose response relationships and hedonic price relationships (if applicable).
- Site and socioeconomic characteristics.
- Comments/summary.
- Related/other studies.

The following considerations are made when selecting studies to be included in ENVALUE (Nash 2004).

- Australian study.
- Policy requirements of the NSW EPA and NSW government, specific interests of staff or foreseen needs.
- Having dollar values or dose-response functions.
- Usefulness for benefit transfer.
- Example of valuation methods of interest to other researchers.

The selection criteria thus include an evaluation of a study's usefulness for benefit transfers. In addition, study quality is assessed by a number of criteria that varies depending on what valuation method was used. The criteria are presented in table B.1.9. The number of criteria is not big, but some of them are demanding because they require detailed knowledge of valuation methods. For example, criterion 1 presupposes that the reviewer has insights good enough to be able to judge whether the environmental good was "carefully" measured or not. Moreover, some criteria (7, 10, 11, 15, 16) deal with problems that are not necessarily reported or discussed in the publication about the valuation study. Hence, the reviewer has in these cases to possess knowledge enough for a critical interpretation of the results of the study. However, whether the other criteria are fulfilled or not should be clear from reading the publication.

Table B.1.9. *Evaluation criteria used in ENVALUE for each valuation method. (ENVALUE's terms for valuation methods have been used. Shades indicate that the criterion is not used for the method in question.)*

Evaluation criteria	DR	RRC	HPA	PE	HPM	TCM	CVM	CCM
1. Was the environmental good carefully measured?						■	■	■
2. Were primary data used to measure economic input?		■	■	■	■	■	■	■
3. Were primary data used to measure economic impact?	■					■	■	■
4. Were results affected by household income?						■	■	■
5. Were results correlated with other factors?						■	■	■
6. Were socioeconomic differences accounted for?		■		■		■		■
7. Were there problems from jointness?	■	■			■	■	■	■
8. Were substitute sites accounted for?	■	■	■	■	■		■	■
9. Percentage of travel time included	■	■	■	■	■		■	■
10. Did respondents act as private agents?	■	■	■	■	■	■		■
11. Did respondents conduct their own 'cost-benefit' analyses?	■	■	■	■	■	■		■
12. Method of expressing preference (ranking, rating, choice)	■	■	■	■	■	■	■	
13. Experimental design (number of attributes and levels, number of replications)	■	■	■	■	■	■	■	
14. Form of survey	■	■	■	■	■	■		
15. Were there biases present?	■	■	■	■	■	■		■
16. Other economic/econometric problems								
17. Survey size								
18. Other								

DR: Dose response approach
RRC: Replacement/repair cost approach
HPA: Household production approach
PE: Preventive expenditure
HPM: Hedonic price method
TCM: Travel cost method
CVM: Contingent valuation method
CCM: Conjoint/choice models

Source: Nash (2004).

B.1.3.3 VALUATION STUDY DATABASE FOR ENVIRONMENTAL CHANGE IN SWEDEN (VALUEBASE^{SWE})

ValueBase^{SWE} is a database including information on about 170 valuation studies concerning environmental change in Sweden. It is available free of charge from www.beijer.kva.se/valuebase.htm. A report including a bibliography and abstracts of valuation studies complements the database (Sundberg and Söderqvist 2004b). The purposes of the database and the report were to contribute to the following:

- Increased opportunities to use results of valuation studies in cost-benefit analyses and other tools for decision-making.
- Increased knowledge of methodological development in valuation methods.
- Avoid unnecessary repetitions of valuation studies.
- Support benefit transfers.
- Facilitate networking among persons interested in economic valuation of the environment.
- Increased opportunities to come to general conclusions about the economic value of environmental change, and to carry out meta-analyses of valuation results.
- Facilitate integration of Swedish valuation results into international databases such as EVRI.

The database includes rather detailed information on the valuation studies which involved collection of primary data. The design of data fields in ValueBase^{SWE} was to a large extent inspired by the data fields in EVRI. The 30 data fields contain the information below. However, only bibliographic data are included for studies using secondary data.

- Bibliographic information.
- Type of study: if primary data or secondary data were used or if the study is a meta study or a review.
- Relation to other studies.
- Valuation method (also details such as question format and payment vehicle in the case of SP methods).
- Study area and study population.
- Type of environmental good/service and environmental asset.
- Extent of environmental change.
- Relation to environmental quality objective.
- Sample information: sample size, sampling procedure, response rate, year of data collection.
- Payment vehicle.
- Economic measures and estimated values.
- Valuation function.

- Used in CBA/policy.
- Remarks.

At present, the database does not include any indications on the quality of the valuation studies, so quality is left to users to assess. In our opinion, some of the studies included in the database are not satisfactory from a quality point of view and are consequently not suitable to use for decision-making. This emphasises the need for a method to assess the quality of the studies.

B.1.3.4 THE NEW ZEALAND NON-MARKET VALUATION DATABASE

The purpose of this database is to collect information about all valuation studies that have been carried out in New Zealand. The database is free of charge to use and includes information on about 100 valuation studies (http://learn.lincoln.ac.nz/markval). The data fields are as follows:

- Data year.
- Object of study.
- Type of study (valuation method).
- Item valued (recreation, pollution, aesthetics, risk, community services, transport, environmental preservation or other).
- Mean value.
- Authors.
- Reviews.
- References.

While there is a field for "reviews", studies are at present not being reviewed and some of the studies included in the database have a poor quality (Kerr 2004).

APPENDIX B2

Additional results related to quality

Table B.2.1. *Evaluation questions for valuation studies using the contingent valuation method.*

Basic knowledge needed to evaluate a contingent valuation study:

A. Background

1. Who was the sponsor of the study, and what interests, if any, does the sponsor have in the provision of the amenity?
2. When were the data gathered? Have there subsequently been any major changes in public opinion which are likely to affect the benefit estimates?

B. Sampling and aggregation procedures

1. What population did the study wish to represent in the sample?
2. What sampling plan was used to draw the sample from the population of interest? Was it probability based? How well was it executed?
3. What were the original sample size, the sampling response rates, and the usable number of respondents whose WTP amounts were employed to estimate the benefits?
4. What were the non-response rates to the valuation questions?
5. What effect did the non-responses have on the benefits estimates?

C. Scenario

In evaluating the scenario, three dimensions should be considered:

1. Whether the hypothetical market makes sense from the standpoint of economic theory.
2. Whether the scenario is relevant to the policy being valued.
3. Whether respondents are likely to understand the scenario.

Some key questions are:

4. How was the amenity described?
5. Could an average person understand the description?
6. What property right was assumed?
7. Were the measures used (e.g. WTP or WTA) appropriate for the property right and meaningful to the respondents?

8. Was the amenity being valued distinguished from related amenities with which respondents might confuse it?
9. What types of benefits (e.g. use or existence values) were likely to be included in the respondents' WTP amounts for the amenity?
10. Was the researcher aware of possible sequencing effects? For instance, if a deliberate sequencing effect was not desired and more than one good or level of provision was being valued, were the respondents informed of what they would be asked before they valued the first improvement? Were respondents given a chance at a later point in the interview to revise their amounts if they wished to do so?
11. Were key scenario elements such as the payment vehicle and probability of provision appropriate to the policy being valued?
12. Were respondents provided with sufficient information to enable them to make an informed decision?
13. Was the description of the amenity accurate?
14. To what extent are the descriptions of the amenity and the changes in the magnitude of its provision relevant for policy use?
15. What provisions were made in the wording of the scenario to ensure that the potential sources of bias from instrument effects were minimized?

D. Survey procedures

1. What method was used to gather the data? If a telephone or a mail survey was used, have the special problems posed by these methods been addressed?
2. What procedures were used to develop and pretest the instrument?
3. How was the survey administered? This information will vary somewhat according to the survey method. Of particular importance are such questions as: How was the survey explained to the respondents? Who was described as the sponsor? Who executed the interviews or conducted the mail survey? What procedures were used to ensure that prevailing standards of survey practice were followed?

E. Data analysis

1. What procedures were used to identify and handle outliers and protest responses?
2. Is sufficient information provided about the cases dropped to permit a judgment about the validity of this procedure?
3. What methods were used to compensate for missing data?
4. If a valuation function is estimated, have alternative specifications been considered?
5. Is the valuation function, if any, robust to violations of the assumptions made in estimating it?
6. Are the data available for independent analysis?

F. Evidence of reliability and validity

1. Is the complete questionnaire available for examination?
2. Was the questionnaire (including introductory material and all materials shown to the respondent) clearly worded throughout?
3. Was the descriptive material presented in a way likely to maintain the respondent's interest?
4. Did the questionnaire contain any material that might lead respondents to place a greater or lesser value on the amenity than would be the case in a genuinely neutral instrument?
5. Did the wording overemphasize the hypothetical nature of the study or the impact it could have on public policy in a way that might lead respondents to give strategic responses?
6. Did the information provided about the amenity include all the characteristics necessary for the valuations to be meaningful?
7. Was any consistent design rule used to make decisions about the sample and scenario design? If yes, what is the implication for the findings?
8. What evidence is there that the respondents understood the questions as intended by the researcher?
9. Does the researcher discuss those response patterns of various groups of respondents which are consistent (or inconsistent) with the respondents' understanding of the scenario?
10. Are the results of a meeting held to debrief the interviewers at the conclusion of the study reported?
11. What evidence is there of the effects of potential biases? Are the results of an experiment built into the design of the survey using, for example, split samples?
12. What are the results of a regression analysis of the WTP amounts on a set of theoretically relevant predictor variables? (This would provide evidence of reliability and validity.)
13. Were sensitivity analyses conducted, and if so, what were their findings? (These will aid in assessment of the findings' stability.)
14. What are the statistical confidence intervals for the WTP estimates, based on sampling variability?
15. Has the role of non-sampling errors been satisfactorily addressed, and have appropriate warnings been provided?

Source: Mitchell and Carson (1989, p. 301–303).

Table B.2.2. *Guidelines of the NOAA Panel for the contingent valuation method.*

The following need to be present in order to assure reliability and usefulness of the information that is obtained from a CVM study.

1. *Sample type and size*

   Probability sampling is essential for a survey used for damage assessment. The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

2. *Minimize non-responses*

   High non-response rates would make the survey results unreliable. Minimizing both sample non-response and item non-response are important. The former is unlikely to be below 20 percent even in very high quality surveys. The latter has also been large in some CVM surveys because of the difficulty of the task respondents are asked to perform.

3. *Personal interviews*

   It is unlikely that reliable estimates of values could be elicited with mail surveys. Mail surveys typically employ lists that cover too small a part of the population (e.g. samples based on telephone directories omit approximately half the US population). In addition, since the content of a mail questionnaire can be reviewed by targeted respondents before deciding to return it, those most interested in a natural resource issue or in one side or the other can make their decision on that basis. It is also impossible using mail surveys to guarantee random selection within households or to confine answering to a single respondent, and it is difficult to control question-order effects. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of costs and centralized supervision.

4. *Pretesting for interviewer effects*

   It is possible that interviewers contribute to “social desirability” bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CVM studies should incorporate experiments that assess interviewer effects.

5. *Reporting*

   Every report of a CVM study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g. refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g. advance letters). All data from the study should be archived and made available to interested parties.

6. *Careful pretesting of a CVM questionnaire*

   It is necessary to have very careful pilot work and pretesting plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

7. *Conservative design*

   When aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate WTP is preferred.

8. *Elicitation format*

   The WTP format should be used instead of WTA because the former is the conservative choice.

9. *Referendum format*

   The valuation question should be posed as a vote on a referendum. Asking respondents to give a dollar valuation in response to an open-ended question presents them with an extremely difficult task. At the same time, presenting respondents a set of monetary amounts for which they are to choose is likely to create anchoring and other forms of bias. We recommend as the most desirable form of elicitation the use of a dichotomous question that asks respondents to vote for or against a particular level of taxation. Such a question form also has advantage in terms of incentive compatibility.

10. *Accurate description of the program or policy*

    Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

11. *Pretesting of photographs*

    The effects of photographs on subjects must be carefully explored. The dramatic nature of a photograph may have much more emotional impact than the rest of the questionnaire.

12. *Reminder of undamaged substitute commodities*

    Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resources. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

13. *Adequate time lapse from the accident*

    The survey must be conducted at a time sufficiently distant from the date of the environmental insult that the respondents regard the scenario of complete restoration as plausible.

14. *Temporal averaging*

    Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points of time.

15. *"No-answer" option*

    A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation question. Respondents who choose the "no-answer" option should be asked non-directively to explain their choice. Answers should be carefully coded to show the types of responses.

16. *Yes/no follow-ups*

    Yes and no responses should be followed up by the open-ended question "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses.

17. *Cross-tabulations*

    The survey should include a variety of other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of WTP broken down by these categories. Among the items that would be helpful in interpreting the responses are: Income, prior knowledge of the site, prior interest in the site (visitation rates), attitudes toward the environment, attitudes toward big business, distance to the site, understanding of the task, belief in the scenarios, ability/willingness to perform the task.

18. *Checks on understanding and acceptance*

    The questionnaire should attempt at the end to determine the degree to which respondents accept as true the descriptions given and assertions made prior to the valuation question.

19. *Alternative expenditure possibilities*

    Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario. It is not at all clear how exhaustive should be the list of alternative public goods that are explicitly presented. If the list is too brief, overspending can be expected. If the list is too long, respondents will be encouraged to spread expenditures to public goods for which there is not adequate total demand and which therefore cannot really be offered to them. The survey should probably include some statement about the price of alternatives, for example, the per capita expenditure that would be required to provide the items.

20. *Deflection of transaction value*

    The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated.

21. *Steady state or interim losses*

    It should be made apparent that respondents can distinguish interim from steady-state losses. The quality of any natural resource varies daily and seasonally around some "equilibrium" or "steady state" level. Active-use value of a resource depends on its actual state at the time of use. But passive-use value of a natural resource may derive only or mostly from its steady state and not from its day-to-day state. If so, full restoration at some future date eliminates or greatly reduces passive-use loss.

22. *Present value calculations of interim losses*

    It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

23. *Advance approval*

    Since the design of a CVM survey can have a substantial effect on the responses, it is desirable that – if possible – critical features be preapproved by both sides in a legal action.

24. *Burden of proof*

    Until such time as there is a set of reliable reference surveys, survey designers must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. A CVM survey should be judged as "unreliable" if it suffered from any of the following maladies: (i) a high non-response rate to the entire survey instrument or to the valuation question, (ii) inadequate responsiveness to the scope of the environmental insult, (iii) lack of understanding of the task by the respondents, (iv) lack of belief in the full restoration scenario, and (v) "yes" or "no" votes on the hypothetical referendum that are not followed up or explained by making reference to the cost and/or the value of the program.

Source: Arrow et al. (1993).

Table B.2.3. *Crucial questions to pose for assessing the validity of a CVM study.*

1. Was the true value clearly and correctly defined?
2. Were the environmental attributes relevant to potential subjects fully identified?
3. Were the potential effects of the intervention on environmental attributes and other economic parameters adequately documented and communicated?
4. Were respondents aware of their budget constraints and of the existence and status of environmental and other substitutes?
5. Was the context for valuation fully specified and incentive compatible?
6. Did survey participants accept the scenario? Did they believe the scenario?
7. How adequate and complete were survey questions other than those designed to elicit values?
8. Was the survey mode appropriate?
9. Were qualitative research procedures, pretests, and pilots sufficient to find and remedy identifiable flaws in the survey instrument and associated materials?
10. Given study objectives, how adequate were the procedures used to choose study subjects, assign them to treatments (if applicable), and encourage high response rates?
11. Was the econometric analysis adequate?
12. How adequate are the written materials from the study?

Source: Bishop (2003).

Table B.2.4. *Key factors influencing the size of value estimates made by stated preferences methods.*

- Date of study
- Policy context
- Scale of environmental change examined
- Type of study
- Payment vehicle
- Treatment of protest votes and analysis of outliers
- Consideration of income constraints
- Embedding (part-whole bias)
- WTP vs. WTA

Source: Bann et al. (2003).

Table B.2.5. *A survey of desirable contents in a report on the application of an SP method, according to Bateman et al. (2002).*

A. Objectives

1. A detailed account of the non-market effect being valued.
2. Description of attributes of the non-market effect that might vary in a final programme or policy.
3. Other relevant information concerning the attitudes or opinions of the population that might usefully be collected as part of a survey.

B. Methodology

1. An explanation of the relevant SP technique.
2. Justification for the choice of technique.
3. Interpretation of the expected results in the context of economic theory.

C. Literature review

A comprehensive review of existing valuation studies similar to the current study. Aspects of relevance for comparison between earlier studies and the current study include:

1. The methodology used for valuation; studies using revealed preferences as well as existing meta-studies should be included.
2. Relevant characteristics of the resource or change considered in each case.
3. The country and site of interest.
4. The population sampled (e.g. users, non-users, nationals, non-nationals).
5. Relevant information on the choice of scenario, payment vehicle and institutional context used in the questionnaire.
6. Valuation results: at a minimum, mean and median WTP/WTA should be given for each group, although other aspects (for example, the valuation function) may also be relevant.
7. Evidence of the population affected, and an estimate of the total value of the change.
8. The difficulties and lessons to be taken on in the current study should be highlighted.

D. Population and sampling strategy

At a minimum, the following should be covered in the report:

1. Choice of sample frame population (e.g. visitors) and the reason for this choice.
2. Choice of sample (e.g. quota or probability sample) and the reason for this choice.
3. Choice of survey mode (e.g. in-person interviews, mail survey) and the reason for this choice.
4. The sample size.

E. Questionnaire design

A brief overview of the questionnaire(s), outlining the relevant sections (e.g. attitudes, uses, valuation scenario, socio-economic characteristics) and the objectives of each. For each section of the questionnaire(s), the following should be discussed.

1. The type of data collected, and why it is of interest.
2. The structure of the questions and the techniques used (e.g. paired comparisons, Likert scales).
3. The relevance of the questions (e.g. to help explain WTP answers, to encourage the respondent to think about the relevant issues). For choice modelling studies, description of choices, attributes and attribute levels should also be presented.
4. The structure of the valuation question, including the hypothetical scenario, the payment mechanism and the elicitation technique.

F. Implementation

Pre-survey and main survey findings should be summarised and include:

1. Who conducted the focus group or survey.
2. Timing and location of the focus groups or survey.
3. Field dates and location for the main survey and major pilots.
4. Brief characteristics of respondents and sample size.
5. Main pre-survey findings and how they affected the final questionnaire design.

G. Results: summary statistics

1. Main summary statistics for socio-economic characteristics: number or percentage of respondents with each characteristic of interest.
2. An assessment of the representativeness of the sample compared to the population of interest where relevant (e.g. national or regional households) or profile of the relevant group (e.g. users).
3. Main summary statistics of uses and attitudes: number or percentages of respondents indicating each possible response for each question.
4. Disaggregation according to readily identifiable groups of interest (e.g. users, non-users).
5. Exploration of relationships between variables of interest (e.g. correlation between attitudinal and use variables).

H. Results: analysis of WTP/WTA data

Main findings of the econometric results should be presented in the body of the report, including:

1. Type of data (e.g. WTP/WTA, continuous, binary, interval).
2. Treatment of refusals and protest bids, and checks for any systematic bias in the characteristics of the sample if these bids are excluded.
3. Weighting procedures to correct for lack of representativeness, if relevant.
4. Treatment of missing data (e.g. for income).

5. Specification of the model (e.g. bid function, utility difference model).
6. Model estimation and results including goodness- of fit estimates, including standard errors, t-statistics, (pseudo)R^2 and tests for IIA in conditional logit models.
7. Estimation of mean/median.

I. Validity testing

The study should consider the implications of the following validity tests:

1. Content/face validity testing: whether the study asked the right questions in a clear, understandable, sensible and appropriate manner should be discussed. Findings from focus groups, pilot and main surveys are useful here. Whether there are indications of the existence of scope, embedding and other biases, the likely reasons for these and how they are tackled (if possible) should also be discussed.
2. Convergent validity assessing: whether the results of the SP study are comparable to other market and non-market valuation studies should be presented (if possible) by comparisons of the study results and the results of the literature review section.
3. Expectation based validity testing: whether the SP study results are in line with theoretical and intuitive expectations should be addressed. If there are departures from such expectations, these should be explained.

J. Aggregation and implications

This section of the report should include:

1. Which aggregation strategy was used and why.
2. A discussion of forms of bias, whether they occurred and if so, the strategy to deal with them.
3. Assumptions used in the analysis, with a discussion of their possible implications.
4. An estimate of the total value(s) of interest, with sensitivity analysis to test the effect of the main assumptions upon the results.
5. The sources of supplementary data required for aggregation (e.g. estimates of the relevant population).

K. Annexes

At a minimum, the annexes should contain:

1. The full version of the questionnaire(s) used.
2. Any screening instrument used to select respondents.
3. Detailed econometric analysis of the results.

Source: Bateman et al. (2002).

Table B.2.6. *Guidelines for the use of the contingent valuation method; an example from the UK.*

Minimum requirements for a CVM or choice experiment study

1. Sample size should be a minimum of 500 individuals.
2. This shall be a random sample of the population to which it is intended to generalise the results.
3. Professional quality fieldwork is required, complying with the Interviewer Quality Control Scheme and the Code of Conduct of the Association of Market Survey Organisations. This fieldwork should normally be undertaken by a specialist field-work organisation.
4. Personal interviews with respondents are required rather than postal surveys or telephone interviews.
5. The results may not be generalised to a wider geographical population than that included in the sample.
6. The effects of distance on both the probability that an individual is prepared to pay and the amount that an individual is prepared to pay shall be analysed.
7. The use value component of preparedness to pay shall be removed from preparedness to pay when estimating existence value.
8. Both the likelihood that an individual is prepared to pay and the amount that an individual is prepared to pay shall be reported separately.
9. Since theory predicts that both the likelihood an individual is prepared to pay and the amount that such an individual is prepared to pay depend upon a number of factors, notably income, the extent to which the results are consistent with theory shall be reported.
10. A report of the reasons why the values obtained can be treated as specific to the site in question is required.

Source: Defra (2000, p. 102).

Table B.2.7. *Guidelines for the use of stated preferences methods; an example from Australia.*

Survey design

1. Pre-survey consultation with focus groups comprising a cross section of people who are representative of the population that will be sampled. The role of the focus group is to assist with: (i) defining the attributes, (ii) checking communication aspects of the questionnaire, (iii) checking that the scenarios are plausible and understood, (iv) ensuring that the payment vehicle is appropriate.
2. Pilot test of the survey instrument using a randomly selected sample from the population of interest.
3. Questions to collect attitudinal, demographic and socio-economic information. Attitudinal information is useful because (i) it can be used to check the validity of valuation results by cross tabulating respondent attitudes against the value estimates, and (ii) it can be incorporated as explanatory variables into the stated preference model. Demographic and socio-economic information is required as an input into the modelling phase and it is also useful for checking how well the sample represents the population of interest. At a minimum, data should be collected on age, income, sex, educational status and occupation.
4. Use of follow-up questions in the questionnaire for picking up response aberrations such as (i) payment vehicle protests, (ii) protests that constitute free riding behaviour, and (iii) lexicographic preferences (e.g. options that include an improvement in safety are always chosen irrespective of the cost).

Survey logistics

1. For other than simple CVM techniques or preliminary attitudinal testing where telephone surveys are appropriate, questionnaires to be administered by one of the following methods: (i) mail out/mail back, (ii) personal drop off and pick up, (iii) personal interview, or (iv) centrally administration of the questionnaire where respondents meet at a central location and complete the survey on computer terminals or using pen and paper.
2. The sample size must allow reliable statistical estimation for all choice sets and population segments. A minimum of 10 respondents is required per choice set.

Technical considerations

1. Substitution effects could be important depending on what quality is subject to valuation. For example, WTA is likely to exceed WTP for changes in water quality attributes that affect human health because health has no substitute. In these circumstances, it is recommended that a WTP measure is used as this is likely to produce a conservative, lower-bound value estimate.
2. Embedding exists if individual attributes are evaluated separately and the sum of these values exceeds the value given for the whole package of attributes. Studies should report on the embedding effect and provide a full explanation as to the context in which the quality improvements were embedded or 'framed'.
3. Block design may be used for overcoming the problem of complex designs of choice sets when the number of attributes is relatively large.

Reporting

1. Detailed evidence that any purported differences in WTP between strata in a population are not the result of chance. The simplest approach is to define segments within the sample using distinguishing individual characteristics such as income, occupation, or whether the respondent is a household or an individual. The individual-specific variables are then interacted with various attributes of the choices to produce a model that is specific for a given segment of the sample. Alternatively, more sophisticated methods of accounting for respondent heterogeneity can be used, e.g. random parameters models that account for heterogeneity by allowing model parameters to vary randomly over individuals.
2. Discussion on any evidence of the embedding effect and a detailed explanation of the frame of reference in which quality improvements were presented to respondents.
3. Distinction between marginal value and total value.
4. Full reporting of the methodology and accompanying analysis to derive the estimates.

Source: CIE (2001).

APPENDIX C

A short glossary

Outliers	Observations that differ very much from other observations. For example, a respondent who reported a very high (or very low) willingness to pay in comparison to the WTP of all other respondents might constitute an outlier. Textbooks in statistics provide guidance in defining and identifying outliers.
Explanatory power	The proportion of variation in a dependent variable that can be explained by variations in the explanatory variables. The explanatory power in a regression analysis using the least-squares method is usually measured by (adjusted) R^2.
Nonsampling error	Ín the case of a survey: Other errors than the errors arising because a sample is studied instead of a whole population. For example, nonsampling error might arise because of the collection and processing of data.
Target population	The population that the study actually wants to come to conclusions about.
Unit nonresponse	All values are missing for the object in question, e.g. when an individual has not at all answered a mail questionnaire.
Item nonresponse	Only some values are missing for the object in question, e.g. when an individual has not answered some of the questions in a mail questionnaire.
Frame population	The population that in practice was used in a study.
Probability sample	A sample selected by a method based on a random process, i.e. by a method involving knowledge of the probability of any object in the population being selected to the sample.
Sampling error	In the case of a survey: Errors arising because a sample is studied instead of a whole population.
Sampling frame	The target population constitutes the sampling frame when a sample is drawn for the survey.

Do you want to assess the quality of a valuation study? Or do you need assistance in designing a valuation study? This report provides an instrument that will help you with these tasks.

In recent years, there has been an increasing demand for the inclusion of both benefits and costs in assessments of environmental policy proposals. However, difficulties in estimating the benefits side suggest that the positive effects of environmental policy measures risk being underestimated. One solution to this problem is to launch new valuation studies to increase the knowledge base in areas where few or no studies have been carried out to date. However, this requires a significant amount of time and financial resources. It is therefore important to use results from existing studies to the greatest possible extent.

The purpose of this report is to provide an instrument that enables government agencies and consultancies to make consistent and clear assessments of the quality of existing valuation studies. The quality criteria in the report can also be of help in the design of new studies. We expect the instrument will help to improve the quality of economic analyses and thus provide a sound basis for environmental policy decisions.

ISBN 91-620-1252-5
www.naturvardsverket.se

NATUR VÅRDS VERKET
SWEDISH ENVIRONMENTAL PROTECTION AGENCY