DTLR
TRANSPORT
LOCAL GOVERNMENT
REGIONS

Economic Valuation with Stated Preference Techniques

Summary Guide

David Pearce and Ece Özdemiroglu *et al.*

March 2002

Department for Transport, Local Government and the Regions: London

Department for Transport, Local Government and the Regions
Eland House
Bressenden Place
London SW1E 5DU
Telephone 020 7944 3000
Web site www.dtlr.gov.uk

Further copies of this publication are available from:

Department for Transport, Local Government and the Regions
Publications Sales Centre
Cambertown House
Goldthorpe Industrial Estate
Goldthorpe
Rotherham S63 9BL
Tel: 01709 891318
Fax: 01709 881673

or online via the Department's web site.

ISBN 1 85112569 8

Printed in Great Britain on material containing 100% post-consumer waste (text) and
75% post-consumer waste and 25% ECF pulp (cover).

March 2002

Product Code 01 SCSG 1158

Acknowledgements
This report was prepared by David Pearce and Ece Özdemiroglu from work by a team consisting of Ian Bateman, Richard T. Carson, Brett Day, Michael Hanemann, Nick Hanley, Tannis Hett, Michael Jones-Lee, Graham Loomes, Susana Mourato, Ece Özdemiroglu, David W. Pearce, OBE, Robert Sugden and John Swanson. It includes editorial work by Andrew Gibbons for DTLR.

The team's full affiliations are as follows:
Ian Bateman, Professor of Environmental Economics, School of Environmental Sciences and Senior Research Fellow, Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia and University College London
Richard T. Carson, Professor of Economics, University of California, San Diego, Research Director for International Environmental Policy, UC Institute for Global Conflict and Cooperation and a Senior Fellow, San Diego Supercomputer Center
Brett Day, Senior Research Fellow, CSERGE, University of East Anglia
Michael Hanemann, Chancellor's Professor of Agricultural and Resource Economics and Public Policy, University of California at Berkeley
Nick Hanley, Professor of Environmental Economics, University of Glasgow
Tannis Hett, worked as a Consultant for Economics for the Environment Consultancy Ltd (EFTEC), UK. She is currently an Independent Environmental Economics Consultant based in Canada
Michael Jones-Lee, Professor of Economics, University of Newcastle-upon-Tyne
Graham Loomes, Professor of Economic Behaviour and Decision Theory, University of East Anglia
Susana Mourato, Lecturer in Environmental Economics, Department of Environmental Science and Technology, Imperial College of Science, Technology and Medicine, London
Ece Özdemiroglu, Director, Economics for the Environment Consultancy Ltd (EFTEC), UK
David W. Pearce, OBE, Professor of Environmental Economics, University College London and Honorary Professor, Department of Environmental Science and Technology, Imperial College London
Robert Sugden, holds the Leverhulme Personal Research Professorship in Economics, Department of Economics, University of East Anglia
John Swanson, Steer Davis Gleave, an independent transport planning consultancy based in London

Foreword

Sound appraisal is at the heart of good policymaking, and robust valuation of impacts in money terms helps decision makers to take proper account of them. Some of the costs and benefits of policy decisions can be readily valued because they impact directly on markets. But some cannot, and measures therefore have to be constructed or obtained from complementary markets. It will never be possible in practice to value all impacts, but we should aim to extend valuation to as many as we can. Valuation is implicit in most policy decisions, and it is preferable to make it explicit where possible to improve quality and transparency, whatever objections some may have.

Of the various valuation techniques available, stated preference (SP) techniques are being used to an increasing extent. Indeed they are the only kind of technique suitable in many circumstances. This volume, together with its companion manual,[1] was commissioned by the Department of the Environment, Transport and the Regions in order to increase the accessibility of SP techniques and to set out what the literature tells us about best practice. We hope that they will both promote the use of SP techniques and, most importantly, improve the quality of studies that are undertaken. There are many methodological and procedural pitfalls in implementing such techniques which, if not avoided, can easily discredit the results, thus not only failing to help the policymaker but also bringing the techniques into disrepute.

The material we have put together is addressed to two main groups of people. This document is targeted mainly at those in the policy community, including people who may need to commission or manage valuation studies, who have to understand the nature and quality of the results in order to inform decisions. Policy managers need to be able to challenge work commissioned from practitioners and to ensure that it can withstand critical examination when required to support policy. This cannot be done well without some knowledge of the techniques involved. The companion manual will be most helpful for students and those who themselves undertake SP studies. They need to be aware of latest developments and the requirements of good practice.

Stated preference valuation is a rapidly developing field, and this book provides up-to-date information about the latest techniques and approaches. It has been prepared by a top class team of experts, led by Professor David Pearce and including many of the leading figures in stated preference on both sides of the Atlantic. I'm very grateful to all those involved in getting this material into the public domain.

Appraisal can be contentious and difficult. It is not solely the preserve of economists, although good appraisal is largely underpinned by analysis of costs and benefits. Cost benefit analysis aims for a degree of objectivity and impartiality when comparing the relative merits of feasible options. Nevertheless, the potentially all-embracing nature of good appraisal means that many disciplines have contributions to make, both in the social sciences as part of the decision process and in the natural sciences by providing the

1 I. Bateman *et al*., *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar, Cheltenham, 2002.

necessary evidence as inputs for decisions. Cost benefit analysis reaches its limits when some significant impacts cannot in practice be given money values, and in such cases other techniques may help, such as multi-criteria decision analysis (see, for example, the guidance at www.dtlr.gov.uk/about/multicriteria/index.htm).[2]

My aim is to promote actively the use of valuation where it is a practical proposition. Valuation techniques are frequently used for measuring environmental impacts. They also have a long history of use in the health and safety field, where willingness to pay studies of changes to risk of illness, injury or death have informed many policy and investment decisions. In transport they are used for valuing time savings, and hence congestion costs. I hope that by making stated preference valuation techniques more accessible, this guide and the accompanying manual will enable more policy areas to take advantage of them, using best practice techniques.

Chris Riley
Chief Economist
Department for Transport, Local Government and the Regions

2 *Multi-Criteria Analysis: A Manual*, DETR, London, 2000.

CONTENTS

Introduction

Scope of the summary guide

01 This volume summarises the essential steps required for conducting high quality stated preference valuation studies. It is written for people who may be involved in commissioning or managing stated preference (SP) based appraisals or assessing such work. It will help those who need to decide whether or how to find money values for measurable but unpriced impacts such as environmental effects in order to use these values in cost benefit analysis of policy or project options.

02 The contents have been organised so that readers can access the particular material they need. Parts 1 and 2 set the scene, as follows:

- **Part 1** explains the nature and significance of economic values (Chapters 1–3)
- **Part 2** describes the techniques available for estimating them (Chapters 4–6).

It is not essential to read Part 1, but Part 2 contains key background which is relevant to all valuation studies.

03 Parts 3 and 4 describe the separate stages which are involved in undertaking stated preference work and provide the detailed information needed as a study progresses:

- **Part 3** explains how to choose the survey method, the population and the sample (Chapters 7–10), and then discusses how to handle piloting/testing and revision of the questionnaire (Chapter 11)
- **Part 4** covers analysis of the data (Chapters 12–14), and aggregating and reporting the results (Chapter 15).

04 Some people — such as consultants and academics — will need more detailed treatment of particular issues which are covered or mentioned in this summary guide. A companion volume ('the Manual') is being published for DTLR and the material discussed in this summary guide is supported by more detailed explanations in the book.[3]

3 I. Bateman *et al*., *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar, Cheltenham, 2002.

Background

05 Estimating money values for goods and services which don't normally have prices is important for making many decisions, not only those involving public expenditure (see Chapters 2 and 3). Even if such valuations are not explicit, decisions may still involve the use of implicit values. This summary guide explains how money values can be determined by stated preference methods, which are the most comprehensive and commonly used set of techniques.

06 Stated preference techniques rely on asking people hypothetical questions, rather like a market research interview. The aim is to see how people respond to a range of choices, and thus to establish the extent of collective *willingness to pay* for a particular benefit (or their *willingness to accept* payment in exchange for bearing a particular loss). Stated preference questionnaire-based techniques can be contrasted with *revealed preference* analysis which aims to deduce people's willingness to pay from observed evidence of how they behave in the face of real choices.

07 Stated preference is based on what people say rather than what they do, but it is more flexible than revealed preference and can potentially be applied in almost any valuation context. Hypothetical *payment scenarios* can be defined in great detail in order to produce conclusions about people's willingness to pay for either specific aspects or the entirety of goods, services or other things that are relevant to the decision.

08 Stated preference valuation techniques are complex and time consuming, and so need to be undertaken by specialists (see Chapter 4). Using a sound methodology is crucial for the credibility of results. Analysis can be exposed to detailed and hostile scrutiny, especially if it is intended as a basis for policy proposals, and inadequate approaches or skimpy data will be exposed. It is therefore essential that the steps outlined in this guide are followed, to provide the firmest possible foundation for public policy.

09 This summary guide shows the key elements of what the techniques entail and includes the latest developments in this rapidly changing field. It should be useful to the non-specialist and anyone who needs to understand the results of stated preference work. It also summarises the essential steps involved in conducting stated preference studies, and gives pointers to how SP work should be approached and the issues that need to be addressed in order to get reliable and credible outputs. Where further detail is appropriate, references are given to chapters in the Manual.[4]

4 I. Bateman *et al*. (2002), *op. cit*.

The appraisal framework for decisions

10 Decision making in central government in the UK is based on a general appraisal framework which involves the assessment of costs and benefits and associated risks. The full implications of relevant options are examined and compared in terms of their estimated impact on general welfare. There is a presumption that market prices will normally reflect social values and so can often be used to derive welfare effects. In cases where market prices clearly do not reflect collective values (for example, environmental and other effects for which there is no direct market), then *shadow prices* should be estimated. Stated preference valuation techniques are a way of doing this.

11 The elements of money value are examined in Chapters 2 and 3. While some commentators object to putting money values on environmental or other unpriced assets, the alternative is to risk that things which people care about will be not given adequate recognition when decisions are made. If these issues are omitted from decision making, there is a strong risk that non-marketed goods will be under-supplied in the economy, and that non-marketed bads will be over-supplied. Deciding how much of a good to supply, or how much of a bad to tolerate or abate, requires that the value of those goods and bads be brought into balance with the costs of providing the good (or the cost of reducing the bad). In this regard at least, 'money counts' because prices provide an indicator of preferences.

12 Appraisals which are undertaken to support decision making (apart from the purely financial) fall into three broad categories:

- **cost benefit analysis:** where all the significant pros and cons of a range of alternative solutions are compared, ideally in money terms;
- **cost effectiveness analysis:** where alternative ways to meet a defined result are compared, again generally in terms of money value costs;
- **multi-criteria approaches:** which compare alternative options on the basis of attributes which are measured but not necessarily valued.

13 Money values are central to the first two of these approaches. So management decisions may be needed about whether to commit resources to a stated preference exercise in order to generate money values. The potential costs and duration of such studies are discussed in Chapter 4. The third approach inevitably involves some element of weighting, even if only implicitly, so valuations should still be an important component where possible.

The steps in estimating money values

14 Reliably estimated money values for unpriced environmental or other effects will reflect people's willingness to pay for (or accept) certain changes.

15 SP techniques for estimating values are costly and time consuming, and if another alternative is practicable is should be utilised first. Taking values from one context to use in another — *benefits transfer* — can avoid the cost and time penalties inherent in original research. Approaches that might work in a particular context include: transferring willingness to pay (WTP) values from a previous study; using WTP values derived from a meta-analysis of several studies; or transferring a WTP function from another study. The necessary conditions for successfully using these approaches are outlined in Chapter 6. The decision context, the urgency and resources available may enable the policy manager to decide whether there is scope for using benefits transfer.

16 Within the class of SP methods, there are two alternative groups of techniques: choice modelling (CM) and contingent valuation (CV).[5] The techniques for estimating money values are outlined in Chapter 4. Contingent valuation concentrates on the non-market good or service as a whole, while choice modelling seeks people's preferences for the individual characteristics or attributes of these goods and services. Selecting the most suitable stated preference technique is an issue on which expert advice should be sought, as the decision should be informed by a fairly detailed understanding of the techniques and their characteristics. Advice may be available from in-house specialists, but often a recommendation will be needed from specialist consultants, perhaps as part of a tender process for carrying out a valuation exercise.

17 Choosing a valuation technique is dealt with in Chapter 5. It explains that CM is suitable for finding willingness to pay for or accept changes in *characteristics* of the item in question. CV is more appropriate if there is a clear need for analysis based on the whole good rather than on its constituent attributes, or for valuing a sequence of items of projects comprising a larger programme. It may often be sensible to start with a presumption for CM, unless it appears that the process disadvantages of CM would be avoided by using CV. In some contexts it may be sensible to use more than one technique in order to test results for consistency.

18 Once a SP technique has been chosen, decisions are needed on survey method, sampling and initial questionnaire design. The main survey options are face-to-face, telephone interviews or mail surveys, along with various combinations. The pros and cons of these are set out in Chapter 7, noting in particular the benefits (and the higher costs) of face-to-face surveys. The process of estimating willingness to pay is done with respect to a defined target population. The sample will be drawn from this, and sampling processes need to be chosen with regard to appropriate sample size, for which cost constraints may pose a trade-off against precision of estimate. Chapter 8 explains that reliable estimates require statistically adequate sample sizes, and there is no way round this.

19 Questionnaire design for contingent valuation and choice modelling is summarised in Chapters 9 and 10. The CV approach entails asking respondents direct questions about their maximum or minimum willingness to pay for a good or service. The context is a hypothetical but plausible *scenario* which includes a description of the item in question and the proposed *payment vehicle* (such as a tax or a donation). CM focuses on rankings or ratings, which may be easier for respondents to deal with, although money values are

5 Further variants within these exist, and are explained in detail in Bateman *et al. op. cit.*

introduced into each choice option in order to provide a common yardstick. For example, choosing which car offers the best value for money in terms of its features may be easier than deciding how much extra you would be willing to pay for more of a particular feature like faster acceleration or better fuel economy. For both CV and CM, various approaches are discussed to formulating questions in order to get the most accurate responses and to avoid the problems of unrealistic, biased or strategic answers.

20 Piloting questionnaires, revising questions and repeating this sequence is crucially important to ensure that questions and the responses they elicit are as good as possible. Unsound questions or problematic responses may well undermine the credibility of the whole venture — or may allow others to do so once the results are published.

21 Part 4 deals with analysing the sample data gathered at the survey stage, statistically testing its reliability and validity, and reporting the results. Sound econometric analysis of the data is an essential requirement for obtaining outputs which accurately reflect the population's willingness to pay. That level of detail is not covered in this summary guide. It should be undertaken by professional econometricians. The analysis stage should estimate the mean and median willingness to pay (or accept) of respondents, determine the extent to which differences in responses can be explained by respondents' characteristics, and provide a transfer equation for use in future benefits transfer exercises. The end results of the analysis stage are estimates of welfare changes from the proposed scenarios. The tasks this involves are outlined in Chapter 12 for contingent valuation, and in Chapter 13 for choice modelling. Tests for the results' reliability and various aspects of their validity (generally common for both CV and CM approaches) are explained in Chapter 14.

22 The final set of tasks is aggregating and reporting the results. This involves moving from the estimate of WTP for the sample to that for the relevant population. Chapter 15 explains how it may be necessary to apply strategies at the aggregation stage of estimating the bid function in order to compensate for shortcomings which may have emerged in the work of earlier stages (for example, unrepresentative samples or failure to define the target population adequately). Documenting and reporting of research results should be comprehensive, both to facilitate examination and replication of the approach and to provide for possible use of benefits transfer to other contexts which are as yet unknown. A checklist is provided, giving the main areas for which information should be provided.

The aims of the summary guide

This guide is a self-contained document, which summarises the essential steps involved in conducting stated preference studies. However, the guide is simply that — a guide — and it cannot substitute for a detailed manual on how to carry out stated preference studies, the pitfalls, theoretical background and applications. In particular, the processing of stated preference results involves complex statistical procedures and these cannot be adequately summarised in a guidance document. This guide is designed to give those who deal with the results of stated preference studies, or who wish to monitor generally how they are conducted, a clear idea of what is involved. The reader intending to carry out stated preference studies for themselves should consult the companion Manual, which provides comprehensive and detailed guidance.

PART 1

Explaining economic value

CHAPTER 1
What is economic valuation?

Summary

This chapter explains the key economic concepts relevant to economic valuation, including markets, cost benefit analysis and appraisal, the role of human wellbeing, preferences and the components of economic value. The chapter introduces the main stated preference valuation techniques of contingent valuation and choice modelling.

1.01 This guidance document explains how to assign **economic values** to goods and services that do not usually have a market price. It is concerned with **non-marketed goods and services**, and their counterpart **'bads'** — such as pollution, risk and disamenity. The basic rationale for assigning economic values is that if they are omitted from appraisals, there is a strong risk that non-marketed goods will be under-supplied in the economy, and an equally strong risk that non-marketed bads will be over-supplied. Deciding how much of a good to supply, or how much of a bad to tolerate or abate, requires that the value of those goods and bads be brought into balance with the costs of providing the good (or the cost of reducing the bad). This balancing exercise involves some form of **policy and project appraisal**. One form of appraisal that makes full use of economic values is **cost-benefit analysis (CBA)**. But economic valuation has other uses besides cost-benefit analysis. More detail of the rationale for assigning economic values in CBA is provided in Chapter 2. These rationales and the use of valuation cannot be understood without first understanding the theoretical basis for economic valuation.

1.02 **Economic valuation** refers to the assignment of money values to non-marketed assets, goods and services, where the money values have a particular and precise meaning. Non-marketed goods and services refer to those which may not be directly bought and sold in the market place. Examples of non-marketed services would be clean air and neighbourhood amenity. Accident prevention would be an example of a service which is sometimes marketed (purchase of safety equipment) and sometimes not (good driving behaviour).

1.03 If a good or service contributes positively to **human wellbeing,** it has economic value. Whether something contributes to an individual's wellbeing is determined by whether or not it satisfies that individual's **preferences**. An individual's wellbeing is said to be higher in situation B than in situation A if the individual prefers B to A. The basic value judgement underlying economic valuation is that **'preferences count'**, although this does not imply that all decisions must be made on the basis of what people want. Other factors, such as what is morally right, what is just and fair now and for the future, and what is administratively feasible, must be taken into account, although such issues are less susceptible to formal analysis.

1.04 Preferences are revealed in many ways, but the context of interest is the **market**. In the market place, preferences show up through individuals' **willingness to pay (WTP)** for the good in question. A disservice, or 'bad', has negative economic value — it detracts from human wellbeing. The economic value of 'bads' will show up as a willingness to pay to avoid the bad in question, or as a **willingness to accept (WTA) compensation** to tolerate

the bad. WTP in the market place is made up of two components: what is actually paid (the **price**) and the excess of WTP over the price — **consumer's surplus**. Consumer's surplus is therefore a measure of the net gain from the purchase of a marketed good. In a pure non-market context, all WTP is consumer's surplus because there is no market price. In practice, some expenditures will often be incurred in securing the non-market good (e.g. the cost of travelling to a wilderness area), so the consumer's surplus is again the net gain.

1.05 Broadly, there are two ways of estimating the economic values attached to non-marketed goods and services (and bads): using **revealed preferences** or **stated preferences**. Revealed preference approaches identify the ways in which a non-marketed good influences actual markets for some other good, i.e. value is revealed through a **complementary** (surrogate or proxy) **market.** An example of a revealed preference approach would be the measurement of the economic value of noise nuisance as reflected in house prices: houses in noisy areas are likely to be cheaper than comparable houses in quieter but otherwise similar areas. Stated preference approaches on the other hand are based on **constructed markets**, i.e. they ask people what economic value they attach to those goods and services. In other words, the economic value is revealed through a hypothetical or constructed market based on questionnaires. The third approach to economic valuation relies on the build-up of case studies from revealed and stated preference studies and then seeks to 'borrow' the resulting economic values and apply them to a new context. This is **benefits transfer**.

1.06 This guide is concerned with the stated preference approach for eliciting economic values. While terminology can vary, a useful broad distinction within stated preference is between

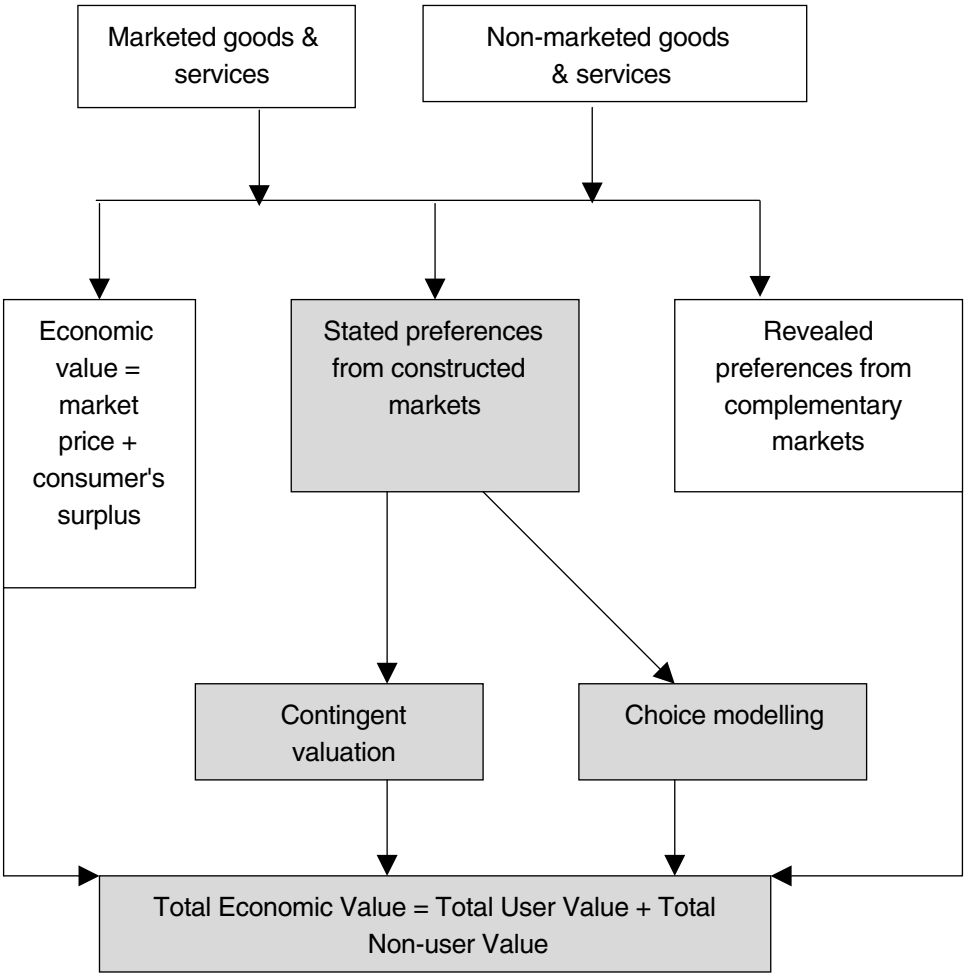

Figure 1.1 The structure of economic valuation

contingent valuation and **choice modelling**. In contingent valuation, respondents to a questionnaire are asked directly for their WTP ('What are you willing to pay?' or 'Are you willing to pay £*X*?'). Choice modelling refers to a variety of procedures for inferring WTP from sets of rankings or ratings of alternative options presented to respondents.

1.07 The aim of economic valuation techniques is to uncover the **total economic value** (TEV) of the good in question. TEV identifies all the changes in human wellbeing that accrue from a change in the provision of the good. Those values may accrue to **users** — persons who make direct or indirect use of the good — and to **non-users** — persons who are willing to pay for the change in the provision of the good but who make no direct use of the good. An example would be willingness to pay for the conservation of endangered species, even though the individual may not have seen, nor expect to see, the species in question.

1.08 Figure 1.1 summarises the main concepts introduced so far. Shaded boxes show the techniques covered by this guidance.

Further reading

The theory underpinning economic valuation is welfare economics. See A. Boardman, D. Greenberg, A. Vining and D. Weimar, *Cost-Benefit Analysis: Concepts and Practice*, Prentice-Hall, New Jersey, 1996.

CHAPTER 2
Reasons for economic valuation

Summary

Economic values are needed for a variety of purposes and underpin a range of different decisions. This summary guidance is principally concerned with the broad context of appraisal issues, but money values are also used in other contexts, often to do with specifying or auditing performance, notably with regard to environmental issues. Besides cost benefit analysis, the contexts in which valuation is used may include approaches to 'green accounting' and sustainable development issues. This chapter explains why money values are important for policy and project appraisals where cost benefit analysis (CBA) is the overarching framework, and gives examples of contexts in which particular types of valuation may be relevant.

2.01 Any appraisal requires criteria for choosing between alternatives. Different criteria may entail trade-offs, such as between cost and quality or performance, as shown in the traffic speed example below. Cost benefit analysis uses money values as weights, because they express people's willingness to pay (WTP) or willingness to accept compensation (WTA). This produces the important characteristic that benefits and costs can be directly compared, and specific actions can be compared with doing nothing (i.e. the base case scenario).

2.02 Cost benefit analysis adopts various decision rules. First, a project or policy is potentially acceptable provided benefits are greater than costs. Benefits are measured by WTP to secure the benefits. Costs may comprise WTA compensation for losses, plus resource costs (e.g. costs of inputs such as labour, capital, raw materials). Since market prices also reflect WTP, resource costs are also linked to WTP. CBA is therefore a decision procedure that is fully consistent with the use of WTP and WTA as measures of economic value. Second, where there are alternative options, it is generally the case that the option with the highest ratio of benefits to costs will be preferred (provided it has benefits greater than costs). Third, time can be incorporated into CBA via the use of discount factors which show the rate of trade-off between current and future gains and between current and futures losses. The resulting **net present value** of benefits (less costs) should be positive in order for a proposal to be accepted as potentially worthwhile.

2.03 CBA is not the only use for economic values. A full list of potential uses of economic valuation is shown in Table 2.1. These uses are wider than is often appreciated. Some applications are widely practised, for example the use of economic values to adjust the national income accounts to reflect the depreciation of natural assets and for environmental pollution. Some applications are in principle quite general, but in practice have been confined to a few studies. Thus the UK's use of economic valuation to inform the setting of tax rates on landfill waste and aggregates extraction has been unique. Priority setting within a programme of actions (e.g. environmental policy) is also feasible in principle using benefit-cost ratios. In practice, such exercises tend to be very data-intensive and costly to undertake. The European Commission's deliberations on its 6th Environmental Action Plan have been partly informed by economic valuation exercises designed to secure benefit-cost ratios for action on climate change, acidification, major

accidents and noise nuisance.[6] The World Bank has pioneered the use of economic values in indicators of sustainability, primarily through adjustments to modified national income accounts, but also through attempts to measure the total wealth of economies. A growing use of stated preference techniques is to show how individuals trade-off benefits or costs over time in order to reveal their own discount rate. This has been pioneered mainly in the health care field but has also been applied to risks to life.

2.04 Valuation techniques are often applied to some outcome of a **dose-response function**, a relationship between an event (the 'dose') and its effects (the 'response'). Thus, air pollution may be the dose, and a health effect may be the response. Dose response functions are scientific relationships which are observed or hypothesised, and are not themselves a valuation technique, but they produce an effect that can be valued, in this case the health effect. In some contexts, dose-response relationships may not be directly identifiable or quantifiable (for example, disamenity from road traffic may include a mix of visual intrusion, noise, vibration and adverse local air quality).

Table 2.1 **Applications of economic valuation techniques**

Context	Comment	Type of valuation likely to be relevant
Cost-benefit analysis: projects and programmes	Traditionally, this is the context in which CBA was developed. Usually used for public investment projects in public or quasi-public goods, but can include non-marketed private goods (e.g. patient care)	RP, SP, BT
Cost-benefit analysis: policies, incl. regulations	A more recent focus in the UK but RIA now required for all regulations. Traditional for many RIA requirements in the USA	RP, SP, BT
'Demonstration' of the importance of an issue	Usually used to estimate economic damage from some activity (e.g. ill-health from pollution)	Usually BT only
Setting priorities within a sector plan	Used for prioritising road investments	Usually BT only
Setting priorities across sectors	Rare, but has been used for this purpose by World Bank	Mainly BT
Establishing the basis for an environmental tax or charge	Recent UK experience appears to be unique, e.g. landfill tax, aggregates tax	Mainly BT (landfill tax) but can include original RP and SP (aggregates tax)
'Green' national income accounting	Only utilised in a limited way in the UK	Usually BT only
Corporate green accounting	A few studies exist, but even fewer are public	BT only
Sustainability indicators	World Bank uses 'genuine savings' and 'wealth' indicators, based on economic valuation, to determine if economies are sustainable or not	Tends to be BT only
Legal damage assessment (liability)	Not yet used in the UK but extensively used in the USA in context of liability for damage. Expected forthcoming in EU.	RP, SP and BT
Estimating discount rates	Used in health (and environmental) literature and for estimating discount rates in developing countries	SP

Note: CBA = cost benefit analysis; RIA = Regulatory Impact Assessment; RP = revealed preference; SP = stated preference, and BT = benefits transfer.

6 Available at europe.eu.int/comm/environment/enveco/priority_study/index.htm.

2.05 The procedure most frequently used in the applications shown in the Table is benefits transfer (BT). This technique is generally quicker and cheaper (when it is practicable), but its validity is open to debate since comparatively few tests of the accuracy involved in transferring values have taken place (see Chapter 6).

2.06 Good decision-making involves, at the very least, adopting criteria by which a policy, project or programme is judged to be 'good', and then doing an appraisal of how the alternative options compare. Analysis should begin with a performance matrix (also known as a criteria-alternatives matrix). This provides the basic building block of rational decision-making.

2.07 An example of a performance matrix is shown below in Table 2.2. The context is the appropriate speed limit on a motorway. The alternative policy options (speed limits) are shown in the four right hand columns of the matrix, with 80 km/h being the base case (i.e., current position or *status quo*). The criteria by which the desirability of changes in the speed limits could be judged are shown to the left of the matrix. In this example the criteria are the number of serious accidents, travel time saved and the cost of operating vehicles. The cells of the matrix then show estimates of the effect of the change in speed limits on each of the criteria used. The numbers are only illustrative.

2.08 The second matrix, Table 2.3, uses the same information but shows the data for the options in terms of comparisons to show differences from the base case. Injuries and cost increase with higher speed (denoted by +) but time spent travelling falls (denoted by −), so any decision on higher speed options would require a trade-off to be made between the criteria.

2.09 Tables 2.2 and 2.3 make it clear that the choice of the 'right' speed limit depends on factors over and above the 'basic' information provided about the effects of speed limits. What is required is some mechanism for **trading off** the time, cost and injuries, i.e. we need to know at what rate the benefits of saving time (denoted by +) can be traded for increased cost and increased injuries (denoted by −). We can use peoples' preferences, as revealed

Table 2.2 **Criteria/alternatives matrix: original data**

	Alternative speed limit options km/h			
Criteria	**70**	**80**	**90**	**100**
Serious injuries (number) per million vehicle km	5.0	5.4	5.9	6.5
Time spent travelling (years) per million vehicle km	3.3	3.1	2.9	2.7
Vehicle operating costs (£million) per million vehicle km	12.6	12.8	13.2	13.8

Table 2.3 **Criteria/alternatives matrix: normalised on 80 km/h**

	Alternative speed limit options			
Criteria	**70**	**80**	**90**	**100**
Serious injuries (number) per million vehicle km	−0.4	0	+0.5	+1.1
Time spent travelling (years) per million vehicle km	+0.2	0	−0.2	−0.4
Vehicle operating costs (£million) per million vehicle km	−0.2	0	+0.4	+1.0

Table 2.4 **Criteria/alternatives matrix: monetised outcome**

	Alternative speed limit options			
Criteria	**70**	**80**	**90**	**100**
Money value of serious injuries (£million)	+0.4	0	−0.5	−1.1
Money value of time spent (£million)	−0.2	0	+0.2	+0.4
Money value of vehicle operating costs (£million)	+0.2	0	−0.4	−1.0
Aggregated net benefit (£million)	+0.4	0	−0.7	−1.7

through willingness to pay, as the means of making the trade-off by expressing all the impacts in money terms. To illustrate, Table 2.4 shows what happens if we adopted the following WTP figures: each injury is valued at £1 million and a year of saving time is also valued at £1 million. Then the computation for each option is simple, as the money values for each criterion are added together. The bottom line shows that the highest net benefit would be for a change in the speed limit downwards to 70 km/h.

2.10 Table 2.4 uses WTP as the weighting mechanism to inform the trade-off required for the policy options to be appraised. In theory any set of weights can be used, provided there is a sound rationale. For example, experts might place 'scores' on the various changes, the scores reflecting their judgement about the relative importance of the different impacts of the changes in speed limit. Cost-benefit analysis uses monetary weights which, in turn, derive from WTP (or WTA) measures, and these in turn reflect individuals' preferences.

2.11 Weights should not be arbitrary: there should always be some rationale for choosing one set of weights rather than another. The first point to establish is that weights are necessary: it will be only in rare cases that any one option will dominate others. Dominating means that an option is strictly better on the basis of one of the criteria being used to assess the policy and at least as good on every other criterion. Deciding *not* to choose weights is formally equivalent to weighting everything equally, i.e. choosing unitary weights.

2.12 Choosing between different weighting procedures requires careful consideration of the ultimate rationale for policy. Preference-based weights are better if there is a view that what people want matters. Expert-based weights may be preferred if it is thought that the public are not, and cannot be, informed about the true consequences of their choices. Economic values are one form of preference-based weight. Public opinion polls involving ratings of options could also provide preference-based weights. Individuals' preferences may be informed by considerations other than self-interest, so that preference-based approaches can be consistent with notions of fairness and justice, for example. One advantage of stated preference techniques is that they permit investigation of the motives for preferences.

2.13 The rationale for using WTP and WTA as weights is:

- the benefits of a policy can be compared directly with the costs of a policy. CBA enables the total economic value of the benefits to be compared with the total economic value of the costs of a policy. In Tables 2.2 to 2.4, CBA is able to show which of the speed limit changes is best, and is able to determine whether the status quo should be relaxed at all. Other assessment procedures enable choices to be made

between options, but may not be able to determine whether the status quo should be changed;[7]

- economic values are preference-based and therefore meet an underlying democratic principle;

- WTP and WTA allow the distributional impact of decisions to be taken account of; and

- as indicators of rates of trade-off, economic values can be used to determine trade-offs across time, as well as within time. The discount rate is, in this sense, an economic value.

2.14 The main problems with using WTP and WTA as weights in stated preference work, as with many other forms of survey, are that:

- individuals may be poorly informed about the consequences of choices and may therefore make the wrong choices; and

- the alternatives may be difficult for individuals to perceive and comprehend (e.g. very small changes in the probability of an accident).

A more detailed analysis of the pros and cons of using WTP and WTA is given in Chapters 11 and 14.

2.15 In conclusion, money valuation is a key pillar of CBA as well as for other applications. Economic values expressed in money terms, if properly determined, will reflect people's preferences and can thus be used as weights to inform any policy analysis or decision process.

Further reading:
The different uses of economic valuation in the environmental context are explored in D.W. Pearce, *Economic Values and the Natural World*, Earthscan, London, 1993. Applications to CBA in various sectors are reviewed in G. Walshe and P. Daffern, *Managing Cost-Benefit Analysis*, Macmillan, Basingstoke, 1990. Early experience in the UK public sector is discussed in Department of the Environment, *Environmental Appraisal in Government Departments*, HMSO (now The Stationery Office, London), 1994.

The performance matrix is an important basis for policy analysis. See for example D. MacRae and D. Whittington, *Expert Advice for Policy Choice*, Georgetown University Press, Washington DC, 1997. The impossibility of escaping from some comparison of gains and losses, and hence the impossibility of avoiding weights, was established clearly in H. A. Thomas, 'The animal farm: a mathematical model for the discussion of social standards for the control of the environment', *Quarterly Journal of Economics*, 1963.

7 Consider, for example, an assessment procedure based on costs and lives saved. Policy A costs £1 million and saves 5 lives. Policy B costs £1 million and saves 3 lives. Given the cost, Policy A is clearly better on **cost-effectiveness** grounds. But this procedure cannot answer the question of whether either policy is worthwhile, i.e. whether it is sensible to spend £1 million at all.

CHAPTER 3
Total economic value

Summary

This chapter explains the concept of total economic value (TEV) and its components, building on the ideas relating to value which have already been discussed. There are different types of value that individuals may apply, for a variety of reasons which often relate to how things are encountered or used, and understanding the relevant concepts is essential for addressing any stated preference issue.

3.01 The economic value of something can be regarded as the extent to which people would be prepared to sacrifice something else in order to obtain or safeguard a quantity of it. Total economic value comprises the sum of **use and non-use values.** Use values may be **direct** (e.g. by consuming the good, visiting a site) or **indirect** (e.g. by securing some benefit from the good). A forest, for example, serves both direct and indirect use functions. Visitors to the forest make direct use of it. The role of the forest in protecting the regional watershed would be an example of an indirect use, as would the role of the forest in sequestering carbon dioxide.

3.02 In addition to current use values, individuals may be willing to pay to conserve the option of future use. If the option relates to their own use, this WTP reflects **option value**. If the future use which individuals are willing to pay for is for others (e.g., children or future generations), it is termed a **bequest value**.

3.03 **Non-use values**, also known as passive use values, arise in contexts where an individual is willing to pay for a good even though he or she makes no direct use of it, may not benefit even indirectly from it, and may not plan any future use for themselves or others. This is also referred to as **existence value**.

3.04 In determining the benefits of a given policy, what matters is the TEV of the benefits which are secured. Categorising TEV into types of economic value is useful in that it provides a checklist of impacts and effects that need, in principle, to be valued. For example, it is easy to overlook non-use values in decisions that relate to, say, conservation of the countryside. The economic value of the countryside will be determined not just by the values placed on it by countryside residents and by visitors, but also by, for example, urban people who simply want the countryside to exist. In the same way, the costs of a policy comprise the TEV of those costs.

3.05 The benefit-cost condition which establishes whether a policy is worthwhile is that TEV of benefits should be greater than TEV of costs.

3.06 How complete is the notion of total economic value? Some people identify **intrinsic values** which are necessarily embodied in the object of value and which therefore are not

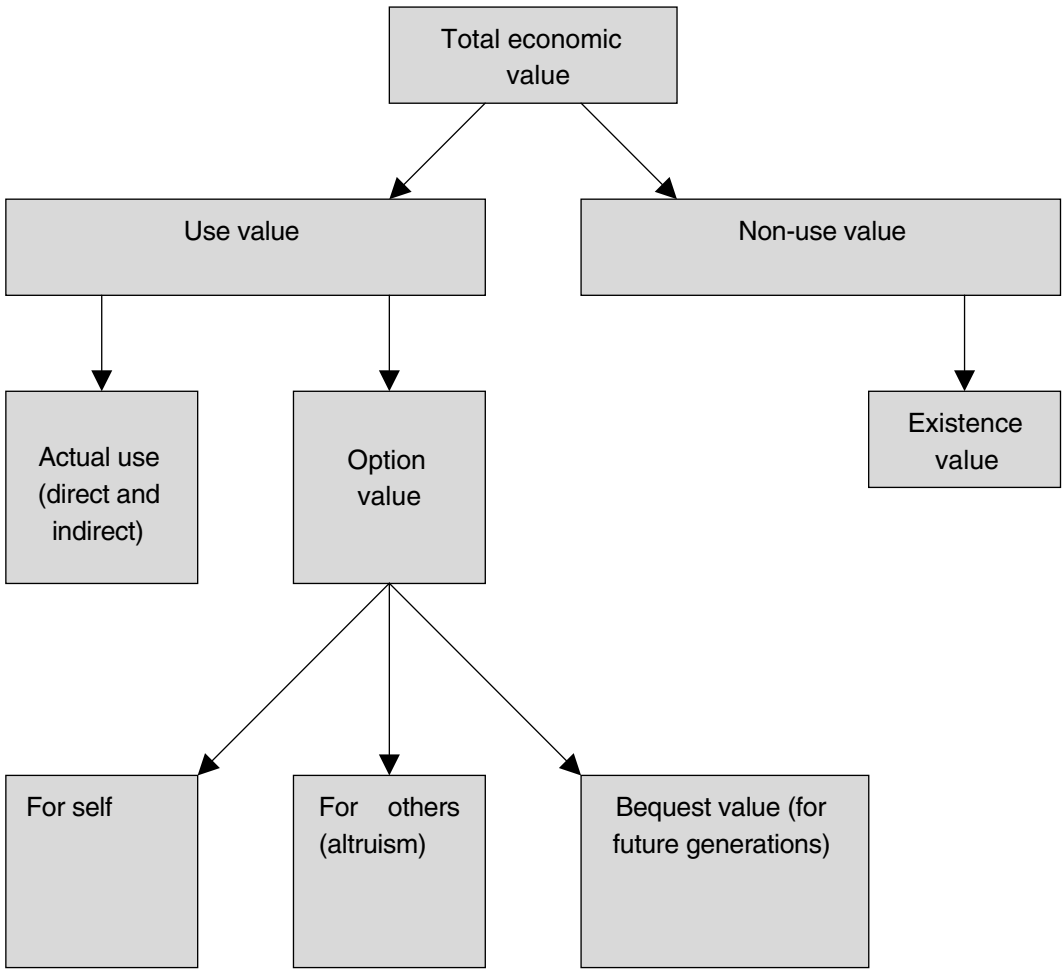

Figure 3.1 Components of total economic value

dependent on human perceptions of value. In the environmental context, this view is often associated with the standpoint of deep ecology. Philosophers dispute the validity of intrinsic value as a concept unrelated to the 'human view': some argue that it is not possible to have views about the status and value of nature that are anything other than anthropocentric. From the practical point of view, intrinsic value, even if it is a valid notion, presents formidable problems for decision-making. If all assets have intrinsic value, and that value is unmeasurable, then no decision could be made that involved a trade-off. But making trade-offs is, as Chapter 2 showed, unavoidable in a world of finite resources.

3.07 Figure 3.1 gives a representation of how the various components of TEV are related.[8]

Further reading

The philosophical debates surrounding the notion of economic value, including an extensive discussion of intrinsic value, are surveyed in W. Beckerman and J. Pasek, *Justice, Posterity, and the Environment*, Oxford: Oxford University Press, 2001.

8 Different authors classify the components differently. Thus, bequest value is often classified as a non-use value because the person expressing the value makes no use of the asset in question. But the bequest is effectively for potential future use and hence is classified under option value here.

PART 2

Estimating economic values

CHAPTER 4
Implementing a stated preference study

Summary

This chapter provides a work programme for anyone who needs to commission a stated preference valuation study, and identifies key considerations in organising the work. Anyone commissioning a stated preference study needs to be familiar with the various stages of design and analysis that such studies require. It is important to understand that failure to go through all these stages could undermine the credibility of the study and thus limit its value as the basis of a decision. SP studies take time to design, pre-test, implement and analyse and hence consultants should always be given adequate time to carry out the study.

4.01 Specific procedures are required if the study is to meet the highest professional standards and to withstand the detailed public scrutiny often applied to work which forms the basis of policy decisions. Consultants must demonstrate that their team contains experts who are:

- familiar with the latest developments in a rapidly growing stated preference literature;
- capable of careful design of questionnaires and surveys, and
- familiar with the econometric procedures required for data analysis.

4.02 To ensure that these conditions are met, and subject to the confidentiality that may surround discussions of such studies, seek external expert advice early on *before* the terms of reference are issued. One or two days' consultancy may avoid major mistakes that will cost many times this amount.

4.03 Ensure that acknowledged independent experts are appointed to membership of the steering group of the study. If, for some reason, independent experts cannot be appointed to the steering group, arrange for a process of peer review. Try to involve peer reviewers from the outset, and well before study drafts are at a stage where the consultant has no real flexibility to change a report in the light of peer reviewers' comments.

4.04 All contracts will be open to competitive tender. You should encourage established groups to register their interest in tendering, and not just rely on those who seek registration, or on past consultants.

4.05 You should inform the tendering consultants of the policy context of the stated preference study. This should include the possible uses to which the study may be put, indications of major disagreements within a department or across departments about the role that SP techniques may play, and the sensitivities concerning any likely outcome of the study. This provides the consultants with a chance to present information in a more helpful way, rather than discovering later that a lot of valuable effort is 'shelved' or invalidated because of issues that could perhaps have been accommodated in the study design.

4.06 The workplan for a standard SP study is presented in Figure 4.1. Both contingent valuation and choice modelling studies follow this structure of tasks, even though they differ in the questionnaire design and data analysis stages. The boxes on the right hand side of the figure include references to the relevant chapters in this guide.

4.07 Always ask contractors for a work schedule and deadlines. However, consultants are probably best placed to know how long a given piece of work will take, and asking for excessively tight deadlines can compromise quality. **The most common weakness of terms of reference is asking for too much in too short a time** and failing to be flexible with respect to consultants' time. Ultimately, if the research can be challenged by those

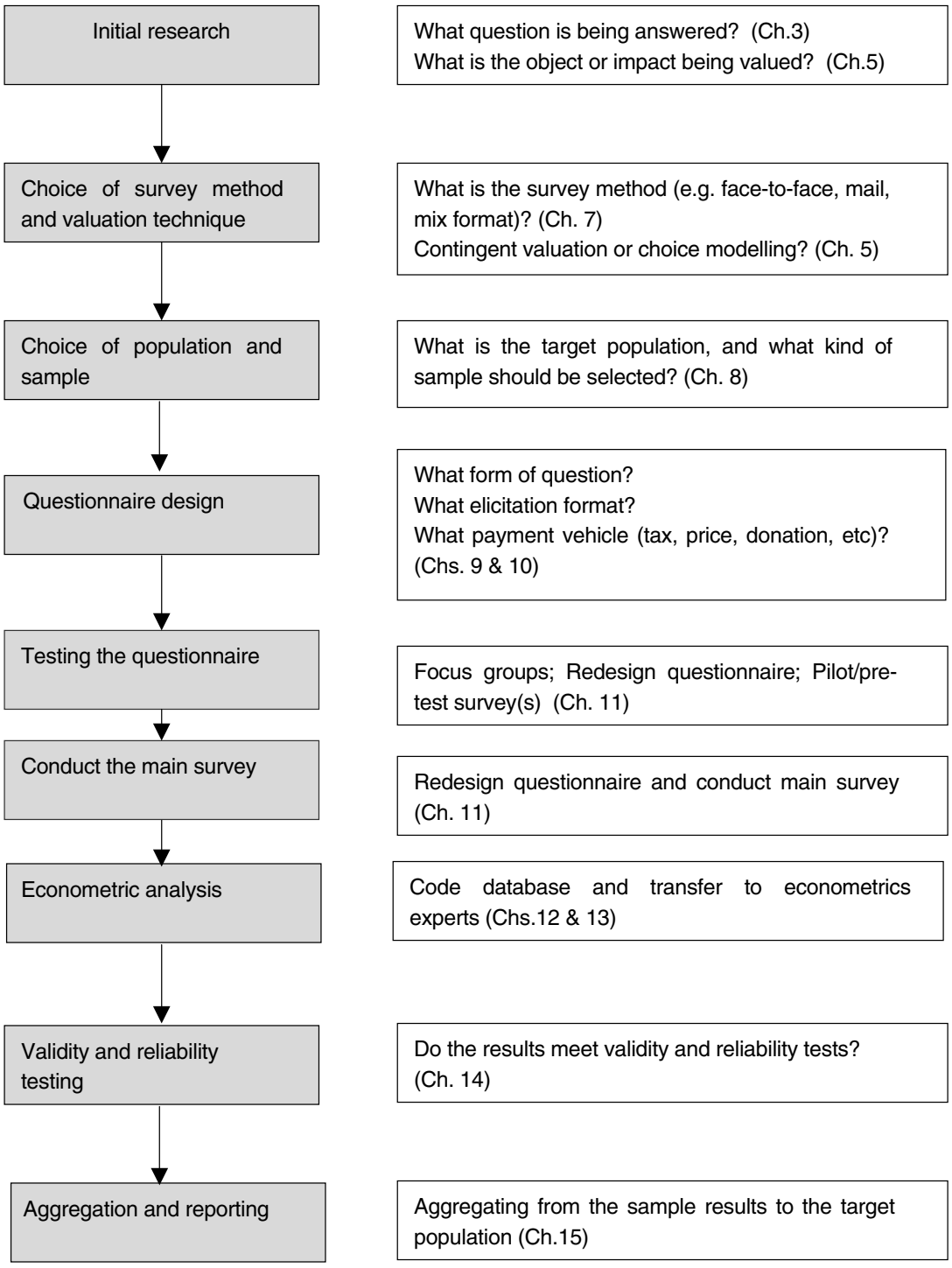

Figure 4.1. The stages of analysis in a stated preference study

stakeholders affected by the outcome, not only is more time taken up, delaying the policy process, but serious damage can be done to the credibility of the techniques being commissioned.

4.08 An additional quality guarantee is to invite consultants to state their own plans for dissemination in the event that the document will be made public. Journal publication, for example, adds to assurances about quality control.

4.09 If time permits, invite the consultants to present the results of the study to a seminar to which known adversaries are also invited. Ensure that actual questionnaires are routinely provided in the report and to seminar participants — this is a requirement for academic publication in many cases.

4.10 Giving guidance on the cost of SP studies is complex. Cost depends on issues such as the number of samples to be surveyed, size of individual samples and their geographical location, the various questions that the survey is designed to answer, and the nature of the survey. As a rough guide, the cost per interview is lowest for mailed interviews (mailshot) at perhaps £10–20 per interview, £15–30 per interview for telephone interviews, and £25–50 for a face-to-face interview (all prices are 2000 values). Note that this is the cost for the survey fieldwork component only. The total cost of a study will also include, at the least, the experts' time for design and testing of the questionnaire and analysis of data, and training of interviewers. In general, studies costing as little as £8–10,000 should be treated with considerable caution. It is unlikely that reliable research for a single sample study can be carried out for less than £25–30,000 (excluding the field survey costs).

4.11 The next two chapters introduce the main methods of determining economic values in money terms for non-marketed effects and explain how they are used.

Further reading

There is an extensive discussion of how to conduct surveys and how to ensure professional results in D. MacCrae and D. Whittington, *Expert Advice for Policy Choice: Analysis and Discourse*, Georgetown University Press, Washington DC, 1996. Although aimed at developing country practice, the advice given in the following paper is generally applicable: D. Whittington, Administering Contingent Valuation Studies in Developing Countries, *World Development*, 26, 1, 21–30, 1998.

An early landmark document containing guidelines for any CV study intended to produce information useful in natural resource damage assessment is Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H. 'Report of the NOAA Panel on Contingent Valuation,' *Federal Register*, vol. 58, no. 10, pp. 4601–4614, 1993, also available at www.darp.noaa.gov/pdf/cvblue.pdf

CHAPTER 5
Choosing between economic valuation techniques

Summary

This chapter first explains some of the techniques available for measuring economic value. It then shows how to choose between revealed and stated preference techniques, and explains which of the various stated preference techniques may be useful in different contexts.

5.01 As we have seen in Chapter 3, total economic value consists of both use and non-use values. Use values can be estimated by both revealed preference and stated preference techniques, but non-use values can only be estimated by stated preference techniques.

5.02 Figure 5.1 shows the relationship between total economic value and the two main categories of valuation technique. Those techniques that are the subject of this guidance are highlighted.

5.03 Valuation may be **context specific** or **context free.** Context specific valuation would be appropriate if the value people apply to a good or service depends on the circumstances and

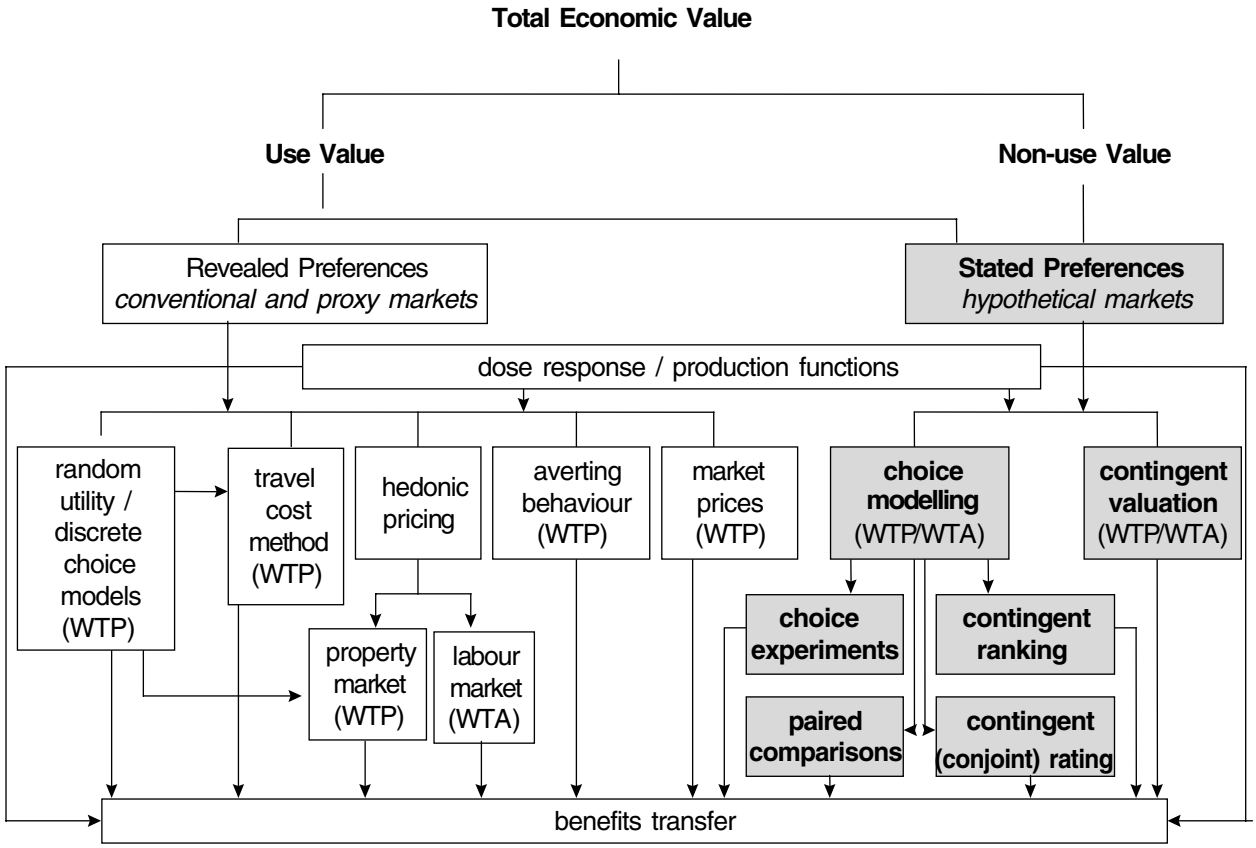

Figure 5.1 Total economic value and valuation techniques

the source of the effect, for example air pollution. Context free valuation would concentrate solely on the item to be valued.

5.04 Impacts can be valued by using various approaches based on **revealed preference**, using direct observation of actual values for complementary effects:

- **Discrete choice models** work on the basis that choices between alternative options reflect the wellbeing (utility) that accrues from those options.

- Choices expressed in probabilistic form, i.e. in terms of the probability of choosing one option rather than another, define the **random utility model** which underlies **travel cost approaches**. Travel cost approaches are used to value recreational assets via the expenditures on travelling to the site.

- **Hedonic pricing** refers to the measurement of effects which show up in labour markets or property markets. Thus occupational risks may be reflected in wage premia, and disamenity effects such as noise may be inferred from house prices.

- **Averting behaviour** involves expenditures to avoid unwanted effects (e.g., smoke alarms, double-glazing, child-proof containers).

- Many effects do show up directly in markets (e.g. the effects of pollution on crops which have a market value).

5.05 Stated preference techniques, asking people hypothetical questions, are classified into **contingent valuation** and **choice modelling** techniques. The former seek measures of willingness to pay through direct questions such as 'What are you willing to pay?' and 'Are you willing to pay £X'? The latter seeks to secure rankings and ratings of alternatives from which WTP can be inferred. The various forms of choice modelling are: choice experiments, contingent ranking, paired comparisons and contingent rating. These are discussed in Chapter 10.

5.06 For any given problem, **which valuation technique should be chosen: revealed preference (RP) or stated preference (SP)?** Both RP and SP can be used to measure use values, but the first issue is to consider whether both options are in practice available. In some contexts there may be no appropriate proxy markets from which to estimate RP values, in which case some form of SP is the only option. If suitable proxy markets are available, then RP approaches can be used as long as they fit the needs of the analysis.

5.07 Revealed preference cannot be used to estimate non-use values (NUV), as shown in Figure 5.1. SP is the only approach that can uncover them. Contingent valuation and choice modelling have both been used to estimate non-use values. Determining in advance whether NUV are likely to be important involves securing expert advice from those familiar with the empirical literature. This literature tends to suggest that NUV will be important if the asset in question is fairly unique and/or has some 'heritage' feature. Examples might be a cultural monument, or an ecologically important forest or wetland. Additionally, NUV is likely to exist where people feel that the asset is important for educational purposes. Examples might be museums and libraries. The empirical literature can be surveyed to see the contexts in which significant NUV have been found, although care has to be taken since many studies do not set out to value NUV. The absence of NUV estimates cannot be taken to mean NUV is not important.

5.08 While SP techniques can, in principle, be used to value any impact, in practice there may be **cognitive limitations** to stating preferences. People may not fully understand, for example, very small changes in risk, or highly complex goods such as biological diversity. Whether or not these limitations exist can be tested by (a) reference to the available literature and (b) use of focus groups. In the event that individuals cannot identify with the good whose provision is being changed (hypothetically, or in reality) then revealed preference techniques will also tend to produce 'wrong' estimates. This is because the 'failure' of perception will translate into a failure of behaviour to respond to the effect in question, and hence market prices will not show any response to the effect.

5.09 SP techniques may nonetheless yield valuable data, especially if sub-samples of respondents can be provided with different levels of information about the good. An advantage of SP is therefore the ability to vary some of the features of the good in question. With RP, the analyst cannot necessarily determine fully what factors lie behind a given valuation. For example, a property price might vary with an environmental variable because of aesthetic concerns, health concerns, concerns about nuisance, and so on (although regression analysis can use dummy variables quite successfully to proxy for such concerns in hedonic price studies). Without surveys, the motives for preferences cannot be discerned.

5.10 SP and RP techniques can be used together. One may be used as a check on the other, a form of convergent validity (See Chapter 14). Analysis of combined studies suggests that contingent valuation produces, on average, WTP values just below those of RP techniques. But RP techniques cannot identify option and non-use values. SP techniques are the most widely used method for valuing non-market impacts in cost benefit analysis. SP will typically be chosen over RP because it provides for the inclusion of non-use values, and the necessary data may not be available for revealed preference techniques.

5.11 The second major question is, if SP is chosen over RP, **which SP technique should be used?** The answer depends in part on how much detail is required on the characteristics of the good or effect being valued. Sometimes it may be desirable to use both CV and CM, to increase the robustness and to check the underlying components of values. Any good can be thought of as a bundle of characteristics, and if the focus of interest is on the economic value of different characteristics of the good in question, then choice modelling techniques are more likely to be relevant. However, changes in the characteristics of interest should not be correlated, and there should be a clear trade-off between them. On the other hand, some studies need to answer questions only about the good as a whole (e.g. what is the economic value of a wetland?). But there may be contexts where what matters are the characteristics of the wetland — the composition of viewing sites, potential for boating or camping, amenities on site, or even the mix of wildlife. These would be relevant for, say, a management decision rather than a decision about whether the wetland should be protected in general.

5.12 In principle, all valuation techniques can be used to value characteristics as well as the whole good (provided NUV is not important — if it is, then RP is not relevant). In practice, choice modelling will generally serve best for this purpose. Contingent valuation can be used to value characteristics, but the questionnaire can quickly become cumbersome and too demanding for respondents. While RP techniques can value different characteristics (e.g. house prices may be 'decomposed' into the value of features of the property itself, neighbourhood characteristics etc.), statistical problems may arise if characteristics are correlated with each other. Choice modelling is designed to avoid these problems by not including options (i.e., choice sets in the survey) in which characteristics are correlated.

5.13 Choice modelling may avoid some of the response difficulties that can be found in contingent valuation. For example, dichotomous choice designs (giving yes/no answers) in contingent valuation may still be subject to 'yea-saying', with respondents giving affirmative but possibly false responses, despite improvements in design standards. Yea-saying is a problem which takes two forms: first, a respondent may try to please the interviewer and say 'yes' when really he should truthfully say 'no'; and secondly, an individual may say 'yes' to a much higher bid than his own valuation if this is the only way he can register a vote, knowing that the higher amount will not be collected from him. Choice experiments avoid this problem, since respondents get many chances in the interview to express a positive preference for a valued good over a range of payment amounts. Open-ended contingent valuation designs (how much are you willing to pay? rather than 'Are you willing to pay £*X*: Yes or No?) avoid the yea-saying problem, but are generally viewed by many experts as facing respondents with a mental task which may be very difficult. Choice experiments face respondents with a much easier problem: Do I prefer A, B or neither?

5.14 In choice modelling, it is necessary to assume that the value of the whole good is equal to the sum of the parts, i.e. its attributes. This may not be valid. In order to test whether it is a valid assumption, 'whole' values from choice modelling must be compared with values obtained for the same resource using some other technique (such as contingent valuation). In the transport field for example, studies show clear evidence that whole bundles of improvements can be valued at less than the sum of the component values.

5.15 To recap: the main difference between CV and CM is that if values for individual characteristics/attributes are required, then CM is preferable. This is because the format of a CV survey required to generate such information would be too cumbersome and long. The specification and treatment of attributes needs care, as the rule with CM is to avoid correlation between characteristics. In order for respondents to face preference-revealing choices, attributes should change differentially (for example, with regard to improvement vs degradation). If all attributes improve or worsen together, there will be no useful information revealed about preferences.

5.16 The extent of analysts' initial knowledge about the potential changes (in ex-ante valuation exercises) is also an important issue. If all aspects of a potential change are known and we are interested in people's WTP to avoid that change (the whole good), for example in the case of a change in a wetland due to an abstraction project, we can then opt for a CV design. However, it is likely that the precise details of the change will not be known, although there will be a general idea of potential impacts on the wetland characteristics such as bird population, water quality etc. Such potential outcomes can be used as levels of the chosen attributes against which preferences can be tested in the survey. Then, after the valuation exercise when the change in wetland is known with more certainty, the decision maker already has information about how people value individual changes — even if they were not shown the whole change.

5.17 Various qualifications need to be considered before finalising the choice of technique. There is no reason to suppose that choice modelling is any better than contingent valuation with respect to hypothetical bias, i.e. actual WTP being less than (and sometimes much less than) stated WTP. Choice modelling may involve greater demands on the patience and concentration of respondents. There is evidence of inconsistent responses that increase as the number of choices increases.

5.18 Contingent valuation is sometimes criticised for insensitivity to scope (known as the problem of 'embedding'), i.e. the situation whereby the value of a good or change is not

significantly different from the value of a more inclusive good or change. For example, 100 units of the good are valued in the same way as 10 or 20 units. Contingent valuation studies seem to pass the scope test, i.e. they tend to find that values are higher for larger 'quantities' of the good. But it is rare to find scope tests which allow one to observe WTP values across a wide range of quantities. Expanding the range of quantities that is offered to the respondent would be desirable from the point of view of estimating whole demand curves, although in practice, financial constraints mean that contingent valuation tends to estimate only those limited parts of a demand curve which are needed for the question in hand.

5.19 Choice modelling might be thought to do better with respect to the range of quantity variations, since scale can be one attribute in the experimental design. Also, the repeated choices used in a choice modelling study mean that more combinations of WTP and quantity can be addressed. Nevertheless, some researchers have found significant scope insensitivity in separate experiments involving choice modelling when the respondents are given too many sets of choices.

5.20 A small percentage of respondents in contingent valuation studies typically refuse to 'play the game' perhaps due to ethical objections to the idea of paying for a given good. Contingent valuation usually treats such responses as protests, and excludes them from the analysis. Choice modelling might serve better here, as it avoids asking for direct monetary valuations of a good, relying instead in statistical techniques to infer WTP indirectly from choices, ranks or ratings. However, this is still an untested proposition.

5.21 Clearly, choosing the right valuation technique is not straightforward. Certain rules of guidance may be discerned however. These are summarised in Table 5.1.

5.22 Table 5.1 does not include benefits transfer, the use of valuations that have already been estimated from other studies. This is the subject of Chapter 6.

Table 5.1 **General guidance on choosing between valuation techniques**

Issue	**Choice of technique**
Non-use value is likely to be important	SP (CV or CM)
Need to maximise credibility of results	SP and RP as checks on each other as far as use value is concerned
Need to value characteristics of the good	CM likely to be preferable, as long as characteristics are not correlated and the proposed changes display trade-offs between characteristics. RP (for use values only) and CV are both possible, but CV likely to handle fewer characteristics than CM.
Markets function freely and flexibly	RP will pick up effects
Motives for valuation matter	RP can decompose values but SP produces detailed attitudinal and motivational analysis
Cognitive load (ability to comprehend nature of change) in doubt	All SP techniques may fail. But if the target effect is unknown to respondents and they thus have no preferences for or against it, then RP is also unlikely to succeed.
Effects of information on valuation need to be known	SP can vary information by sub-sample. RP cannot uncover informational differences between people.

CHAPTER 6
Benefits transfer

Summary

Borrowing suitable values from previous studies and adjusting them to the context in question potentially provides a way to avoid the primary valuation work — as long as the original studies used valid techniques. This chapter explains the principle forms of benefits transfer and the conditions necessary to ensure their validity. Although these are rarely fulfilled, meaning that the technique is seldom ideal, it can sometimes be used when there are practical pressures to short-cut full scale valuation procedures.

What is benefits transfer?

6.01 Benefits transfer (BT) is the process of taking information about economic values (which in fact can be benefits or costs) from one context (the 'study site') and applying it to another context (the 'policy site'). The attraction of BT is obvious: if BT is a valid procedure, then the need for original ('primary') stated or revealed preference studies would be vastly reduced. In the limit, values could be taken 'off the shelf' and applied to new contexts. In fact, this is how a large number of CBAs and other studies in which economic valuation is used have proceeded for some considerable time. But it is only in the last decade that the question has been properly raised as to whether this procedure is valid.

6.02 In benefits transfer, it is possible to (i) transfer an average WTP estimate from one primary study, (ii) transfer WTP estimates from a meta-analysis of several primary studies, and (iii) transfer a WTP function. Each of these procedures is considered in turn.

Methods of transferring benefits

6.03 One elementary procedure is to 'borrow' a single estimate of WTP in context *i* (the study site) and apply it to context *j* (the policy site). The estimate may be left unadjusted, or it may be adjusted in some way. Transferring unadjusted estimates is hazardous, although it is widely practised. Differences in average WTP can arise from differences in:

- the socio-economic characteristics of the relevant populations;
- the physical characteristics of the study and policy site;
- the proposed change in provision between the sites of the good to be valued; and

- the market conditions applying to the sites (for example variation in the availability of substitutes).

6.04 As a general rule, there is little evidence that the conditions for accepting unadjusted value transfer (see paragraph 6.6 below) hold in practice. Effectively, those conditions amount to saying that all the differences listed above do not exist, i.e. sites are effectively 'identical' in all these characteristics. A potentially superior alternative is therefore to adjust the WTP estimates in some way.

6.05 A widely used formula for adjusted transfer is:

$$WTP_j = WTP_i\ (Y_j/Y_i)^e$$

where Y is income per capita, WTP is willingness to pay, and e is the 'income elasticity of WTP', i.e. an estimate of how the percentage change in WTP for the environmental attribute in question varies with a given percentage change in income. In this case, the feature that is changed between the two sites is income, perhaps because it is thought that this is the most important factor resulting in changes in WTP. But it should also be possible to make a similar adjustment for, say, changes in age structure between the two populations, changes in population density, and so on. Making multiple changes of this kind amounts to transferring benefit functions (see below).

6.06 Various forms of adjusted transfer exist in the literature:

- adjustments made by expert judgement;
- re-analysis of existing study samples to identify subsamples of data suitable for transfer; and
- 'meta-analyses' of numbers of previous estimates permitting the estimation of cross-study benefit functions applicable to policy sites.

6.07 Meta-analysis takes the results from a number of studies and analyses them in such a way that the variations in WTP found in those studies can be explained. Meta-analysis proceeds by taking available primary studies and using them as the data to derive an overall explanation for WTP. Factors explaining the variation in WTP in the individual studies would include all the usual factors — income, age, education etc. — but also features of the studies themselves — who carried them out, year of study, size of sample, measures of deviation in average WTP and so on. The resulting meta-analysis equation is then subjected to statistical regression analysis to explain the variability in WTP. This should enable better transfer of values since we can find out what WTP depends on. In the meta-analysis case, whole functions are transferred rather than average values, but the functions do not come from a single study, but from a collection of studies.

6.08 Akin to meta-analysis, but using information from just one primary study, is the transfer of a benefit function from site i to site j. If it is known that $WTP_i = f(A,B,C,Y)$, where A, B, C and Y are factors affecting WTP at site i, then WTP_j can be estimated using the coefficients from this equation but using the values of A, B, C, Y at site j.

Conditions for satisfactory benefits transfer

6.09 There are two broad procedures for validating a BT, i.e. for determining whether the transferred value is likely to be accurate:

- Transfer a value and then carry out a primary study at the policy site as well. Ideally, the transferred value and the primary estimate should be similar. If this exercise is repeated until a significant sample of studies exists in which primary and transferred values are calculated for policy sites, then there would be a justification for assuming that transferred values could be used in the future without the need to validate them with primary studies.

- Take a set of *n* primary studies and use *n*-1 of the studies to estimate the value at the n^{th} site. That 'transferred' value can then be compared with the original primary value at that site.

6.10 Certain conditions should be met for a valid transfer of value to take place. These are widely acknowledged to be:

- the studies included in the analysis must themselves be sound;

- there should be WTP regressions between or within the studies used, i.e. regressions showing how WTP varies with explanatory variables;

- the study and policy sites must be similar in terms of population and population characteristics. Alternatively, differences in population must be accounted for;

- other site characteristics should be the same, or differences should be accounted for;

- the change in the provision of the good being valued at the two sites should be similar; and

- property rights should be the same across the sites.

6.11 Some general findings from the literature are:

- transferring benefit functions is more accurate than transferring average values;

- stated preference studies appear to perform no worse than revealed preference studies in terms of transfer error;

- transfer error using stated preference studies is generally quite large: 1 to 75% if 'outliers' are ignored, up to 450% if they are included;

- there is some reason to suppose that individuals' attitudes are important determinants of WTP in stated preference studies, yet most BT makes little effort to test for variability in attitudes across sites (this suggests that BT would have to be supplemented by social surveys at the policy site);

- meta-analysis of contingent valuation studies can explain a reasonable proportion of the variation in the original studies, but the original studies do not include sufficient

information to test whether more information would have increased the explanatory power of the meta-analysis, and

- the missing information may well be of the motivational type, i.e. why people adopt the value stances they do.

Benefits transfer is not the prime concern of this document. For more detailed information, see the discussion of benefits transfer at the Ecosystem Valuation website: www.ecosystemvaluation.org/benefit_transfer.htm

Conclusion

6.12 BT cannot, at the moment, be extensively relied upon to produce valuation estimates which are statistically indistinguishable from the 'true' values. It may be that values are transferable but that much more information is required before meta-analyses can explain the variation in WTP across studies. On this view, more research will improve the reliability of BT at some stage in the future. Another view is that values are inherently not transferable because what is valued is site specific and because the characteristics of those engaging in valuation are site specific too. At the moment, there is no consensus on these issues. Some studies that have sought to validate transfers have found that the errors involved are of the order of 30–40%, which may well be acceptable for policy purposes since it provides a range of values that could be used for sensitivity analysis. Others have found that primary studies fail to replicate the transferred values. This points the way towards (a) continued reliance on primary studies, (b) conducting those primary studies in a manner that is consistent with future BT tests, and (c) using BT only where there is a reasonable consensus that the errors involved are acceptable.

6.13 The need to improve both the quality of and the scope for benefits transfer underlines the importance of primary valuation studies recording their methodology and results in consistent and accessible formats, for the benefit of potential future users. See Chapter 12 for discussion of the bid function, and Chapter 15.2 for guidance on documenting and reporting a stated preference study.

6.14 Databases containing the results of many different kinds of valuation study are now becoming available. These will permit more extended meta-analysis of studies and should improve the validity of transfers. One international database of values is 'EVRI', the Environmental Valuation Resource Inventory (www.evri.ec.gc.ca). Source documents for UK values (but not the values themselves) are listed in the Environmental Valuation Source List for the UK (www.defra.gov.uk/environment/evslist/index.htm).

PART 3

Designing and testing stated preference questionnaires

CHAPTER 7
Choice of survey method

Summary

Once a valuation technique has been chosen, the next step is to decide the kind of survey to be undertaken. This will depend largely on the resources that are available, including time. In general, quality and reliability require time and money. This chapter compares different data collection methods in the light of quality and cost issues.

7.01 A choice has to be made between the potential survey methods. Considerations in selecting the method of surveying are set out in Table 7.1. Cost and response rates are highlighted to show that the higher the quality of the survey, the higher the cost is likely to be.

7.02 Stated preference practitioners generally recommend face-to-face surveys. Although they are more expensive than other means of surveying the population in question, they permit the use of visual aids and far more flexible questionnaires, have greater potential for controlling the sample so that it is as representative of the general population in question as possible, and yield higher response rates. These advantages do not mean that other forms of survey should never be used. Cost will be a major consideration and only the commissioning client and the consultant can decide on the right balance. However, it is important that those commissioning the study are aware of the very probable loss of information and accuracy that is likely to be associated with other forms of survey.

Table 7.1 **Main data collection methods**

Method	Advantages	Disadvantages
Mail surveys Printed questionnaires are posted to potential respondents	**Relatively inexpensive** Lack of interviewer bias Easier to answer sensitive questions Can be completed at respondent's own pace	**Low response rates 25-50%** Self-selection bias Time-consuming Little control over who fills the questionnaire Fixed question order No clarification or probing possible Restricts the use of visual aids Respondent can alter earlier responses
Telephone interviews Interviewers call potential respondents	Complex questionnaire structures are possible **Cheaper than face to face interviews** Permits probing and clarification Relatively quick to administer Easy to monitor **60-75% response rates**	No use of visual aids Restricts use of lengthy scales Respondent may get tired Respondents may not answer sensitive questions Non-telephone or non-listed respondents not sampled
Face-to-face interviews Interviews take place one-to-one between the interviewer and the respondent either at home or another location relevant to the study (intercept survey)	Highly flexible Complex questions and questionnaire structures are possible Permits probing and clarification Larger quantity of data can be collected Potential for extensive use of visual and demonstration aids **High response rates 70% +** Greatest sample control	**Relatively expensive** Possible interviewer bias Intercept surveys: samples normally not representative and self-selection bias Intercept surveys: questionnaires have to be short
Mixed methods: drop off survey The questionnaire is mailed prior to a visit by the interviewer	Initial personal contact gives survey a 'human face' Shares the advantages of mail and face-to-face methods	Survey form may be lost in interval before calling back **Expensive**
Mixed methods: mail + telephone surveys The questionnaire is mailed prior to a phone call by the interviewer	Gives personal touch to the survey Can complete mailed questionnaire in own time	Shares some of the limitations of mail surveys **Relatively expensive**
Computer assisted interviews Interviewer records responses directly to computer and/or respondent may respond to questions on computer screen	Subsequent analysis is quicker since data inputting stage is not necessary Permits more complex interviews Permits use of e-mail and internet	Possible rejection of 'computer technology' E-mail/internet may preclude random sample unless wide coverage of PCs

CHAPTER 8
Identifying the target population and choosing the sample

Summary

This chapter outlines the sequence of tasks in choosing a sample from a target population. This involves first defining the target population, then finding a sample frame as consistent as possible with the target population, and then selecting a sample from the frame in order to apply the survey. The chapter also explains how sample size can be determined, and how greater precision entails higher cost.

8.01 The survey is applied to a sample of respondents, and the nature of the sample will depend on the target population which it is intended to represent. So if the focus of interest is the general population, the survey will seek a sample of that population. There may, however, be a need to separate out the attitudes and WTP of different **stakeholder groups** (gainers and losers), in which case there should be separate sub-samples to cover these categories. Obvious examples are drivers of cars, pedestrians and public transport users, patients with specified illnesses, people of specified age groups, residents as opposed to visitors, and so on. The full sample must be chosen carefully in order to avoid **biases**. Two main forms of bias must be minimised: **sampling error** — the sample chosen is not representative of the population as a whole, and **non-response error** — some people in the sample frame do not respond, for various reasons. The existence of non-respondents affects how representative the sample is of the population. Chapter 14 discusses this issue further.

8.02 The sequence of steps is:

- Choose the **target population.** This could be the whole population of the UK or of a region, or a defined group of people (e.g. anglers, patients with a given health state, etc.). Defining the target population correctly is a critical step, as getting it wrong can bias the results of a valuation study and thus destroy the credibility of an appraisal. This choice is affected by the good or impact in question and the forms of value that are of interest (non-use values, for example, may require wide geographic coverage). Where the relevant population is not readily apparent, secondary sources may show who would benefit from or pay for the good in question. Possible examples are: tourist board information for studies of open-access goods; angling and boating club membership for water pollution incidents; maps of transmission line routes in respect of valuing associated disamenity; and lists of water rate payers for assessing water improvement benefits. Although access to such data may be restricted, the relevant organisations may be either directly involved in the valuation exercise or have an interest in its findings such that information is made available. It may make sense to have more than one target population if the choice is not clear, as this can provide a range of values in the final results.

- Following the choice of a target population, the next step is to identify the **sample frame population** from which the sample will ultimately be drawn. This should be the

closest practicable approximation to the target population, and might for example consist of all the dwelling units within a city, all the voters registered in a city, all households with a telephone, all the residential and commercial customers of an electric utility, all subscribers to a sport-fishing magazine, all the members of an environmental group, or all the visitors to a particular beach. In many cases the sample frame is an explicit list (e.g. a master list of residential addresses, for example, or of registered voters, or of utility customers, etc). Note that there will be some trade-off between the cost and coverage of a sample frame. Lists of specific, smaller populations may be more readily available than a list of the general population. Typical problems are that the frame may omit certain members of the target population, may include some more than once, or may include people not of interest to the study; and,

- Select a sample from the frame using probability sampling. This selection process involves the use of a randomised procedure, such as a computer-generated list of random numbers. While other sampling procedures are not uncommon, and are acceptable for, say, identifying focus groups, only probability approaches meet the requirements of applying statistical theory to derive the properties of the sample estimators. This makes it possible to correct for bias in sample selection and to construct confidence intervals for the population parameters using the sample data.

8.03 Probability sampling tends to take the forms listed in Table 8.1.

Table 8.1 Types of probability sampling

Form of sampling	Method	Advantages
Simple random	Every element of the sample frame is given an equal chance of being selected.	Simple
Systematic	Select every k^{th} element from a randomly ordered population frame.	Simple
Stratified	Sample frame population is divided into distinct sub-populations, or *strata.* A separate and independent sample is selected for each stratum, using random sampling with either the same sampling fraction for each of the strata (*proportionate stratification*) or different sampling fractions (*disproportionate stratification*). The data are used to develop separate within-stratum estimates. Finally, the separate stratum estimates are combined (weighted) to form an overall estimate for the entire population.	Enables estimates to be derived for each sub-group, even though sub-group may be a small fraction of the population.
Clustered multi-stage	Population is divided into a set of groups or 'clusters' but only a random sample of the clusters is selected. Cluster sampling involves sampling *all* the elements within the selected clusters, but the term is also used to cover multi-stage sampling, in which one selects only *a random sample* of the elements within the selected clusters. An example of cluster sampling would be to divide a city into zones, randomly select a set of zones, and then survey every household within the selected zones. In a multi-stage sample, one would survey only a sample of households within the selected zones.	For surveys of large populations that possess some sort of hierarchical structure, multi-stage sampling is generally more convenient and more economical than one-stage simple random sampling. Multi-stage sampling is attractive when no overall sample frame is available.

8.04 The researcher should consider who is to be selected from within a household. If the aim is to get a household's valuation the person selected must be able to speak for the entire household. Responses may not be random, however. For example, older people and females are more likely to pick up the telephone in telephone surveys, or be found at home in face-to-face surveys, than younger or male household members. In mail surveys, there tends to a preponderance of male as well as older household respondents, and there is also a need to use random selection of adult household members, as this is desirable for good quality surveys. One way to do this in a telephone or face-to-face interview is to ask at the outset for the person 18 years of age or older, say, *who had the most recent birthday.* Another is for the interviewer to ask the initial respondent to list all the household members over 18 by gender and age, and then randomly select a person from this list with whom to continue the interview.

8.05 The choice of sample size entails the familiar trade-off of cost versus precision of estimate. It is determined by the following factors, some of which may require initial judgements prior to the survey:

- **variance in the underlying population** (which shows how varied the data is relative to the average: the bigger the variance, the bigger the sample needs to be);
- the **precision required** in the estimates — the 95% confidence interval is a widely used benchmark;
- the likely response rate;
- the need for any disaggregated information — the more sub-samples, the bigger the overall sample; and,
- the **resources available** for the study.

8.06 Where judgements are called for, professional advice will usually needed from a statistician or survey specialist. Some rules of thumb can be given, and Table 8.2 gives some examples.

8.07 These sample sizes must be expanded if reliable estimates for sub-groups of the population are also needed. Unfortunately, there is no easy solution for obtaining estimates for sub-

Table 8.2 **Examples of appropriate sample size**
Sample size with 95% confidence interval. Sample size is obtained from —

Target population	1,000,000				25,000			
Proportion in true sample	0.5		0.2		0.5		0.2	
Accuracy: standard error x2	+/−3%	+/−5%	+/−3%	+/−5%	+/−3%	+/−5%	+/−3%	+/−5%
Sample size required:	1111	400	711	256	1066	384	682	246
Cost if £25 per interview (£000)	28	10	18	6	27	10	17	6

Note: sample size is estimated from the equation $SE(p)=(pq/N)^{0.5}$ where SE = standard error of a proportion p, p is the proportion of the population with the characteristic of interest, $q=1-p$, and N is the required sample size. In the examples, $p=0.5$ and 0.2. Since the 95% confidence interval can be approximated by sample mean $+/-2\times$SE of the sample mean, the value of SE entering the equation is twice the required level of accuracy shown in the table. Population size determines the 'degrees of freedom' so that 25,000 is taken to mean that 3% and 5% are divided by 2, and 1,000,000 corresponds to dividing by 1.96.

groups. The required sample size increases linearly to a first approximation with the number of sub-groups for which separate parameter estimates are required. If there are two sub-groups, then a sample of about 250–500 *from each* is needed in the event of open-ended contingent valuation, or a sample of about 500–1,000 from each in the event of a closed-ended contingent valuation study. In the face of this, researchers often compromise by accepting a somewhat lower standard of precision for the subgroup estimates than for the overall group estimates.

8.08 Smaller sample sizes can be employed if *more information* per respondent is collected, for example in choice modelling where multiple choices or valuations are elicited from each respondent. However, some caution should be exercised before shrinking the sample size. For both statistical and survey research reasons, there is no scope for a 'free lunch' by reducing the sample size and increasing the amount of data collected from each respondent. The statistical concern stems from the possibility of positive correlation among successive responses from the same subject, which can arise from a variety of sources including fatigue or inattentiveness. To the extent that such correlation occurs, it reduces the amount of *statistical information* in the data obtained from each subject. The survey research concern arises from the fact that eliciting multiple valuations from the same subject is likely to undermine the *realism* of each individual scenario and make it seem more hypothetical than if the respondent were confronted with a single valuation scenario.

8.09 More information about the representativeness of the people sampled will emerge at the stage of analysing the data, which is discussed in Chapters 12 and 13.

CHAPTER 9
Questionnaire design: contingent valuation

Summary

The structure and essential components of a typical CV questionnaire are shown in this chapter, and various alternative approaches to value elicitation questions are explored, along with follow-up questions to help establish the validity of responses. Each component in the questionnaire fulfils an important role. Taken together, they introduce the respondent to the context and relevant background in progressively more detail, and also gather information about the respondent and their understanding of the scenario which are needed to report the results or to establish the validity of the response. For example, some questions may test whether key aspects of the scenario have been understood, while other questions about the degree of familiarity with the good in question will reveal whether the respondent is a user or a non-user. The questionnaire must ensure that three specific conditions are upheld in order to ensure the validity of the results: the non-market good must be carefully defined; the scenario must provide a plausible means of payment; and there must be a plausible mechanism for making the trade-off between consumption of private goods and the good in question.

9.01 Contingent valuation (CV) methodology involves asking a random sample of respondents for their willingness to pay (WTP) for a clearly defined good, or willingness to accept (WTA) a loss. It uses direct elicitation by asking questions that take the form: 'what are you willing to pay?' or 'are you willing to pay £*X*?'

9.02 Figure 9.1 sets out the structure of a typical contingent valuation questionnaire. A choice modelling questionnaire has more or less the same structure but the contents of the 'valuation scenario' are different (See Chapter 10). Each stage is discussed below.

Purpose

9.03 It is essential to state the purpose of the CV questionnaire to ensure that respondents understand the context, are motivated to cooperate, and are able to participate in an informed manner. The context should be as realistic as possible in order to encourage realistic and truthful responses (but not to bias the answers). The interviewers should explain who they are (e.g. conducting a survey on behalf of what organisation), and should assure the respondents that their answers will be confidential.

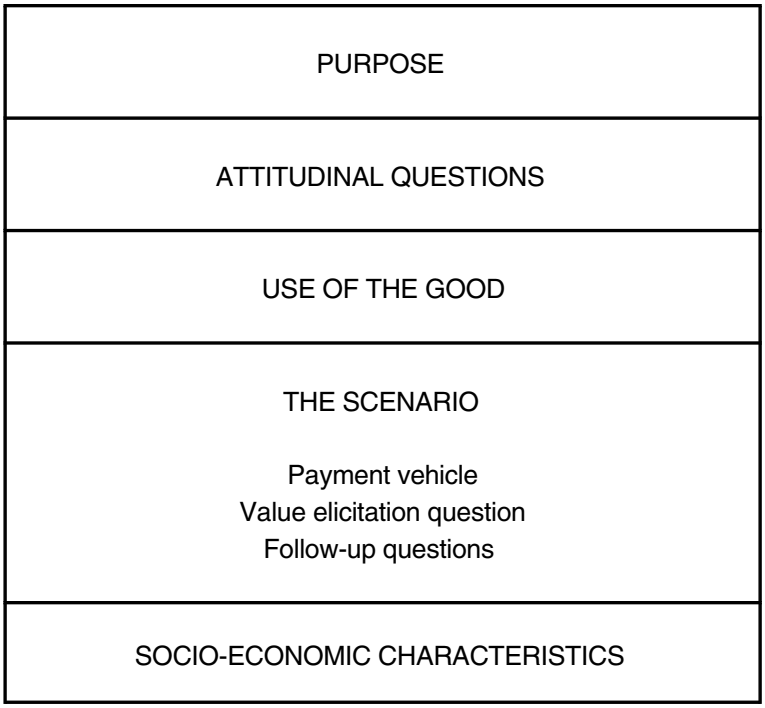

Figure 9.1 Structure of a contingent valuation questionnaire

Attitudinal questions

9.04 The next stage seeks the respondent's attitudes to general issues concerning the good and then to the good in question. For example:

> *Which of the following problems do you regard as being most important for government to solve?*
> *Please tell me with which of the following statements about River X you strongly agree, agree, are indifferent, disagree, or strongly disagree (scale may be shown, e.g. 1 for strongly disagree, 5 for strongly agree)*

Use of the good

9.05 The next stage determines the use of the good or service in question. The aim is to determine what use the respondent makes of the good in order to test their familiarity with it and to distinguish users from non-users. For example:

> *How often do you visit this river?*
> *Which of the following reasons for coming to the river best describes why you come here? (List)*
> *Do you come alone, or with your family, friends?*

The valuation scenario

9.06 The valuation scenario defines the good in question and the nature of the change in the provision of that good. This information makes up a scenario and it is this scenario that respondents will value. Several scenarios may be presented but care has to be taken not to 'overload' respondents so that they become confused about what they are being asked to value. The design of the scenario is a critical feature of a questionnaire: **poorly defined scenarios will elicit meaningless answers.** An example of a well defined scenario is:

> *If no action is taken, this river's quality is expected to deteriorate in the next few years (show picture). To get the river back to its* current state *(show picture) the government will have to spend money and this will mean raising taxes.*

9.07 Note that the scenario **defines the good** (not allowing the river to deteriorate in quality) and the **institution that is responsible for providing the good** (the government in this case). It is important that respondents should have some belief that what they say will influence the decision and that the good will not be provided regardless of what they say. So they must believe the institution in question has the capacity to provide the good. These conditions contribute to the **credibility** of the questionnaire and the scenarios in the questionnaire. It is important, however, to avoid **strategic behaviour**, i.e. the deliberate under- or over-statement of WTP.

The payment vehicle

9.08 The payment vehicle describes the way in which the respondent is (hypothetically) expected to pay for the good. There are no precise rules for choosing between payment vehicles. The nature of the good matters: for example, if it is entirely a local good, a national tax would not be chosen as the vehicle. Table 9.1 lists types of payment vehicle.

Table 9.1 **Types of payment vehicle**

Coercive	**Voluntary**
National tax Local tax Fee/Charge Price increase	Donation to trust fund
Problem: respondent may be hostile to the agency responsible, e.g. hostility to national tax increases may lead to non-response. Test the payment vehicles with a focus group (or a suitable follow-up question in a test sample of surveys)	Problem: voluntary payments invite free-riding. Generally not recommended by CV practitioners.

Eliciting valuations

9.09 The value elicitation question is designed to draw out peoples' willingness to trade goods (or impacts) for money. In this process it is essential to elicit either the **maximum** WTP or the **minimum** WTA in order to be consistent with the underlying theory of economic valuation. Simple WTP or WTA answers are not good enough. This requirement also helps to determine the elicitation format.

9.10 The most widely used elicitation formats are: open-ended, bidding game, payment card, and single-bounded or double-bounded dichotomous choice. **Open-ended elicitation** asks What is your maximum WTP? In a **bidding game** respondents are faced with several rounds of discrete choice questions or bids, with the final question being an open-ended WTP question. **Payment cards** present respondents with a visual aid containing a large number of monetary amounts. Respondents tick sums they are definitely willing to pay and put crosses against those they are definitely not willing to pay. In **single-bounded dichotomous choice** (the 'referendum method') respondents say yes or no to a single WTP amount or bid. With **double-bounded dichotomous choice**, the respondent says yes or no to a stated sum and is then asked to say yes or no to higher/lower bids.

9.11 There is a substantial debate in the literature about the best way to formulate elicitation questions, and the pros and cons of the main formats are shown in Tables 9.2 to 9.6.

9.12 The choice of elicitation format is of considerable importance: different elicitation formats typically produce different estimates. Which elicitation format should be used? Payment cards and dichotomous choice formats can both be recommended. The former is more informative and cheaper to implement than the latter and is superior to both direct open-ended questions and bidding games. The latter may be incentive compatible (encourages truth telling) and facilitates the respondents' valuation task. Both payment cards and dichotomous choice procedures allow for uncertainty in respondents' preferences ('preference imprecision'). The payment card allows for uncertainty across a range of values, i.e. neither a definite 'yes' nor a definite 'no' is appropriate for some range of values. Dichotomous choice procedures can allow for uncertainty by including a 'don't know' option.

9.13 In all approaches, respondents must be reminded of substitute goods as well as the need to trade off money for benefits. Anything which is not unique will have substitutes, which will

Table 9.2 **Open-ended elicitation**

What is the maximum amount that you would be prepared to pay every year[e.g. through a tax surcharge], to improve the landscape around Stonehenge in the ways I have just described?

FOR:	**AGAINST:**
● straightforward; ● does not provide respondents with cues about what the value of the change might be, i.e. no anchoring bias; ● is very informative since maximum WTP can be identified for each respondent; and, ● requires relatively straightforward statistical techniques.	it leads to large non-response rates, protest answers, zero answers and outliers, i.e. unrealistically large bids, and generally to unreliable responses. This is because it may be very difficult for respondents to come up with their true maximum WTP 'out of the blue' for a change they are unfamiliar with and have never thought about valuing before. Moreover, in terms of 'mimicking' markets, most market transactions involve deciding whether or not to buy goods at fixed prices, rather than stating maximum WTP values.

Table 9.3 Bidding game elicitation

Would you pay £5 every year, through a tax surcharge, to improve the landscape around Stonehenge in the ways I have just described?
If Yes: Interviewer keeps increasing the bid until the respondent answers No. Then maximum WTP is elicited.
If No: Interviewer keeps decreasing the bid until respondent answers Yes. Then maximum WTP is elicited.

FOR:
this may facilitate respondents' thought processes and encourage them to consider their preferences carefully.

AGAINST:
- anchoring bias may exist, that is, respondents may be influenced by the starting values and succeeding bids used;
- It also leads to large number of outliers and to 'yea-saying' (giving affirmative but possibly false responses); and,
- Bidding games cannot be used in mail surveys and other self-completed questionnaires.

Table 9.4 Payment card elicitation

Which of the amounts listed below best describes your maximum willingness to pay every year, through a tax surcharge, to improve the landscape around Stonehenge in the ways I have just described? OR *Please place a tick next to the amounts you are certain you would be willing to pay and a cross next to those you are certain you would not be willing to pay. Leave blank those amounts that you are unsure whether you would be willing to pay or not.*

0
£0.5
£1
£2
£3
£4
£5
£7.5
£10
£12.5
£15
£20
£30
£40
£50
£75
£100
£150
£200
> £200

FOR:
- provides a context to the bids, while avoiding starting point bias at the same time (starting point bias being a form of anchoring bias whereby bids are linked to the researcher's statement of the first amount);
- the number of outliers is also reduced in comparison to the previous formats; and,
- some versions of the payment card show how the values in the card relate to actual household expenditures or taxes (benchmarks).

AGAINST:
- vulnerable to biases relating to the range of the numbers used in the card and the location of the benchmarks; and,
- it cannot be used in telephone interviews.

have implications for its value. Respondents must also be reminded of their budget constraints (the limit on their income or wealth) and hence the consequent need to make compensating adjustments in other types of expenditure to accommodate the additional financial transaction implied by the survey.

Table 9.5 Single-bounded dichotomous choice (referendum method)

Would you pay £5 every year, through a tax surcharge, to improve the landscape around Stonehenge in the ways I have just described? (the price is varied randomly across the sample). Yes/No

FOR:

- it is thought to simplify the cognitive task faced by respondents. Respondents have to make a judgement only about a given price, in the same way as they decide whether or not to buy a supermarket good at a certain price, while at the same time providing incentives for the truthful revelation of preferences under certain circumstances (incentive compatibility). That is, it is in the respondent's strategic interest to accept the bid if his WTP is greater or equal than the price asked and to reject otherwise, i.e. ensuring that the respondent tells the truth;
- this procedure minimises non-response and avoids outliers; and,
- the approach received the endorsement of the NOAA panel[8].

AGAINST:

- empirical studies have revealed that values obtained from dichotomous choice elicitation are significantly larger than those resulting from comparable open-ended questions;
- some degree of yea-saying is also possible;
- dichotomous choice formats are relatively inefficient in that less information is available for each respondent (the researcher only knows whether WTP is above or below a certain amount), so that larger samples and stronger statistical assumptions are required. This makes surveys more expensive and their results more sensitive to the statistical assumptions made; and,
- there may also be starting point bias, i.e. answers are 'anchored' on the initial figure stated by the questioner.

Table 9.6 Double-bounded dichotomous choice

Would you pay £5 every year, through a tax surcharge, to improve the landscape around Stonehenge in the ways I have just described? (the price is varied randomly across the sample)
If Yes: *And would you pay £10?*
If No: *And would you pay £1?*

FOR:

more efficient than single-bounded dichotomous choice as more information is elicited about each respondent's WTP. For example, we know that a person's actual value lies between £5 and £10 if she accepted to pay £5 in the first question but rejected £10 in the second.

AGAINST:

all the limitations of the single-bounded procedure still apply. An added problem is the possible loss of incentive compatibility (truth telling) due to the fact that the second question may not be viewed by respondents as being exogenous to the choice situation, and the added possibility of anchoring and yea-saying biases.

Follow-up questions

9.14 It is important to follow-up the answers to WTP or WTA elicitation questions in order to understand the motives behind these answers. Follow-up questions are especially useful where there is some form of protest or unwillingness to pay (or to accept compensation) for the good in question. A protest may show up as an unwillingness to give any answer at all. But **zero valuations are not necessarily protests:** individuals may genuinely not be willing to pay anything for the good. Nonetheless, some zero bids may conceal protest motives. Table 9.7 gives guidance on whether a response should be treated as valid or as a protest bid, in the context of river quality improvements.

8 Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H. 'Report of the NOAA Panel on Contingent Valuation,' *Federal Register*, vol. 58, no. 10, pp. 4601–4614, 1993, also available at www.darp.noaa.gov/pdf/cvblue.pdf

Table 9.7 **Using WTP follow-up questions to determine valid responses**

Possible reasons for unwillingness to pay	**Valid (✓) Protest (×)**
I/our household cannot afford to pay	✓
The change is too small to be of importance	✓
I/we think this problem is not a priority	✓
I am/we would be satisfied with the future situation	✓
I am/we are not very interested in this matter	✓
I do not live near here	✓
There are many other similar goods around	✓
Spending should be on all rivers not just this one	×
I object to paying higher water rates	×
Everyone should pay for this not just local people	×
The government should pay for this	×
The water company should pay for this	×
I need more information/time to answer the question	×
Possible reasons for willingness to pay	**Valid (✓) Bias (×)**
I/we think this problem is important	✓
I /we would like to avoid further deterioration of the river	✓
I am/we are very interested in this river	✓
I /we use this river for recreational purposes	✓
I/we may want to use the river in the future although I/we do not use it now	✓
We should protect the river environment for the animals/plants concerned	✓
We should protect the river environment for future generations	✓
We should protect the river environment for other people to enjoy	✓
I /we get satisfaction from giving to a good cause ('warm glow')	×
I /we will not really have to pay any extra amount	×
My answer reflects my views on the need to preserve all rivers, not just this one	×

9.15 Follow up questions, besides helping to clarify the motives for and validity of responses, can also be used to test the credibility of the scenario. Such questions might ask about the respondent's interest in the good in question, about the need for public consultation, and about the perceived credibility of the institution hypothetically charged with providing the good.

Socio-economic characteristics

9.16 The final section of the questionnaire asks for the socio-economic characteristics of the respondents. This information is used to test whether the WTP answers conform to theoretical expectations (e.g. whether WTP varies with income). A minimum list of such characteristics is age, sex, interests, income (or surrogate measure) and education. Other factors such as nationality and health state may also be relevant according to the issue.

9.17 One possible area of problems with questionnaire approaches to economic value is embedding (and the related issue of scope), whereby the answers given are not sensitive to the quantity of the good being hypothetically provided. The risk of embedding can be reduced by designing questionnaires carefully to avoid poorly defined scenarios and imprecise definitions of the proposed policy change. This and other problems are discussed in Chapter 14. Test questions likely to reveal scope or embedding effects can sometimes be included in the questionnaire if circumstances justify the extra material.

CHAPTER 10
Questionnaire design: choice modelling

Summary

Questionnaires for choice modelling differ from those for contingent valuation (explained in the previous chapter) only in respect of the valuation scenario component. This chapter explains the requirements for the valuation scenario of choice modelling questionnaires and gives examples for the four types of choice modelling: choice experiments, contingent ranking, contingent rating and paired comparisons.

10.01 **Choice modelling** approaches (also known as **conjoint analysis**) are based around the idea that any good can be described in terms of its **attributes**, or **characteristics**, and the levels that these take. For example, a forest can be described in terms of its species diversity, age structure and recreational facilities. Changing attribute levels will essentially result in a different 'good' being produced, and it is on the value of such changes in attributes that choice modelling focuses. Choice modelling (CM) differs from contingent valuation in that it asks for **rankings** or **ratings** rather than for values. CM may therefore avoid some of the problems with **protest votes** since people may find it easier to rank or rate alternatives without having to think in money terms directly. Nonetheless, CM is used to derive economic values through the device of including a money indicator (a price, charge or tax) as a characteristic of each option.

10.02 CM can convey information on which attributes are significant determinants of the values people place on non-market goods; the implied ranking of these attributes amongst the relevant population(s); the value of changing more than one of the attributes at the same time, and the total economic value of a resource or good.

10.03 The types of choice modelling are shown in Table 10.1. The final column asks whether the CM technique is consistent with the underlying theory of welfare economics. As noted in Chapter 1, welfare economics is the basis for economic valuation and for cost-benefit

Table 10.1 **Main choice modelling alternatives**

Approach	Tasks	Estimates consistent with welfare economics?
Choice Experiments	Choose between (usually) two alternatives, versus the status quo	Yes
Contingent Ranking	Rank a series of alternatives	Depends[1]
Contingent Rating	Score alternative scenarios on a scale of 1–10	Doubtful
Paired Comparisons	Score pairs of scenarios on similar scale	Doubtful

1 In order to interpret the results in standard welfare economic terms, one of the options must always be currently feasible.

analysis. The reasons for saying that a particular technique is or is not consistent with the underlying theory are complex, and are not explored here. But note that only choice experiments are unequivocally consistent with the underlying theory of welfare economics. If the context of valuation is, say, cost-benefit analysis, or estimating the size of an externality, then welfare theory-consistent approaches should be used, and choice experiments would be the only possibility.

10.04 The common design stages in CM studies are shown in Table 10.2.

10.05 **Choice experiments** present respondents with a baseline scenario corresponding to the status quo and several alternative options in which specified attributes are changed in quantity. The attributes which are to be included can be determined by focus groups. Chosen attributes should include a money value, which, as in CV, represents a payment vehicle. The number of attributes should be limited to ensure they can be handled by respondents. The choice experiment is constructed as shown below. Respondents are asked to state their choice of A, B or neither. Stating 'neither' means that no benefits are secured relative to the status quo, but no costs are incurred either.

	Option A	**Option B**	**Change in attribute level from A to B (+ better, — worse): illustrative only**
Attribute	A1 A2 A3 A4 (price)	B1 B2 B3 B4 (price)	+ — + +

Table 10.2 **Common design stages for choice modelling**

1. **Selection of attributes**	• Select the relevant attributes of the good to be valued. This is usually done through literature reviews, focus group discussions or direct questioning. Attributes may be chosen because they are the ones most likely to be affected by a policy decision. • A monetary cost is usually one of the attributes to allow the estimation of WTP. Choosing the cost level and the payment vehicle raises the same issues as in CV. • A rule of thumb is not to choose more than 4 or 5 attributes, including cost.
2. **Assignment of levels**	• The attribute levels should be realistic and span the range over which respondents can be expected to have preferences. Levels may include policy targets, and should include the 'do nothing' level and a range about the existing level in order to elicit a WTP for a gain and a WTP to avoid a loss.
3. **Choice of experimental design**	• Statistical design theory is used to combine the levels of the attributes into a number of alternative environmental scenarios or profiles to be presented to respondents. Use of statistical design theory reduces otherwise unwieldy numbers of alternative options.
4. **Construction of choice sets**	• The profiles identified by the experimental design are then grouped into choice sets to be presented to respondents. Profiles can be presented individually, in pairs or in groups, according to the technique being used. Options, i.e. different bundles of attributes, should typically not exceed 8 in number; 4-6 is probably a better working rule.
5. **Measurement of preferences**	• Choice of survey procedure, and conduct of survey. The issues here are common to those met in CV.

10.06 **Contingent ranking** proceeds in the same way as a choice experiment but asks respondents to **rank options** in terms of desirability. In the case below, three options are shown but more than three may be provided.

	Option A	**Option B**	**Option C**
Attribute	A1 A2 A3 A4 (price)	B1 B2 B3 B4 (price)	C1 C2 C3 C4 (price)
Ranking of options: 1.....2.....3.....			

10.07 **Contingent rating** proceeds by drawing up an option as a scenario and asking the respondent to give it a 'rating' on a scale (say, 1...10). The same respondent is then presented with a different scenario and asked to rate that.

Attribute	Option A A1 A2 A3 A4
Tick one level showing your preference for Option A 1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10 *Very low preference* *Very high preference*	

10.08 **Pairwise comparisons** proceed as with choice experiments but respondents indicate their strength of preference for their choice.

	Option A	**Option B**	**Change in attribute level from A to B (+ better, — worse): illustrative only**
Attribute	A1 A2 A3 A4 (price)	B1 B2 B3 B4 (price)	+ — + +
Tick one level 1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10 *Strongly prefer A* *Strongly prefer B*			

10.09 The design and testing of questionnaires should be closely connected, and iteration between the two may be needed. It is important to ensure that design issues can be tried out on small groups first to see whether they give the desired results, and that what is learnt from the testing can be fed back to improve the questionnaire design. Only when questionnaires have performed satisfactorily in focus groups and pilot surveys should a commitment be made to the expense of the full survey. Some aspects of testing are explained in Chapter 11.

CHAPTER 11
Testing the questionnaire design

Summary

Careful questionnaire design is crucial to the success of contingent valuation and choice modelling. This chapter explains the rules for good design and adequate piloting and revision of questionnaires before the full survey. These approaches will avoid the problems of bias that harm the credibility of stated preference valuations. The chapter explains the many kinds of bias that can occur, their main characteristics, and potential solutions. Checks are provided to achieve the goal of high content validity, with a questionnaire that asks the right questions clearly, understandably and appropriately, before the survey is administered.

11.01 Many studies can be criticised precisely because inadequate effort is spent on designing and testing the questionnaire. Questionnaires need to be tested and, if necessary, revised and retested as many times as necessary. A suggested sequence is shown below.

- A focus group is a discussion conducted by a moderator among a small group of respondents. In the context of assessment of non-market effects, focus groups can be used to test the draft questionnaire (e.g. whether the issues are understood, wording is clear, questions are sensible), and to gain insights about issues of interest, particularly methodological issues. Focus groups generally comprise 6–12 people and last for an hour or more. Responses from some participants may be inhibited if more than 12 people are present. Alternatively, a focus group comprising less than 6 participants may encourage single individuals to dominate. Focus groups are, by definition, not expected to be random samples of the population. The questionnaire is often revised in light of the focus group discussion. The number of focus groups required depends on the issue and the questionnaire design, but there should at least one focus group from each sub-sample of interest. After focus group sessions, the questionnaire design should be revised in the light of responses, to eliminate any problems and maximise the amount of information that can be gathered.

- After focus groups, the questionnaire is pilot-tested. The pilot sample may number 25–100 depending on the full sample size envisaged and the complexity of design. It should have some regard for randomness in selection. The questionnaire is often revised again after each pilot, and the number of iterations depends on how much is learnt in each round.

- The final version of the questionnaire is applied to the **full sample**. The desirable size of the full sample was discussed in Chapter 8. Typical full sample sizes are 250–500 for open-ended elicitation formats and 500–1000 for close ended formats (dichotomous choice, payment cards). If the samples need to be split to take account of different valuation scenarios or population groups, full sample sizes need to be increased.

- Depending on the issue being explored, it is possible to have subsets of the main survey which test a certain methodological issue, to discover any differences.

11.02 The work of conducting pilot and main surveys is usually subcontracted to a professional survey organisation, and they can often provide valuable advice on designing or improving questionnaires. Each testing stage should be lead by appropriately trained personnel. If interviewers are not experienced in implementing SP questionnaires, their training should be included as a separate stage of the study.

11.03 As with any commercial relationship, realistic advance planning is needed. Setting this up will often be the responsibility of the principal contractor for the main stated preference study as a whole. The survey company should be brought into discussions with the main contractor sufficiently early to provide advice at the formative stages of the project and to do the pilot and main survey work at the appropriate times. It is sensible for the client to confirm that the contractor plans to use a suitable survey organisation and will involve them in discussions sufficiently early, as survey firms can make valuable comments on questionnaire design on the basis of their previous experience of what works with interviewers and respondents.

11.04 Many pitfalls threaten to undermine the credibility of stated preference results. Careful questionnaire design, piloting and revision can avoid potential biases. These biases are summarised below in Table 11.1. Although the table uses the shorthand term 'WTP', the arguments also apply to WTA.

11.05 Assessing whether or not such biases exist (establishing the question of 'content validity' as referred to in Chapter 14) should not be a retrospective process. It is important that the issues listed above are considered at all stages of the development and administration of the questionnaire. To help achieve this, peer review of questionnaire design, survey administration and data analysis is strongly recommended.

11.06 A questionnaire should aim for high content validity, in which the survey descriptions and questions are clear, reasonable and unbiased such that respondents are put in a frame of mind that motivates them to answer seriously, thoughtfully, and, of course, truthfully. The guiding principle should be whether the diversity of components which constitute a CV study (the sampling frame, administration of the survey, the information and materials provided to respondents, the description of the good and associated provision change scenario, etc.) are conducive and sufficient to induce respondents to reveal valid stated values.

11.07 Once piloting is complete, decisions have been taken about the survey method(s), and the design of the questionnaire(s) has been finalised, then the main survey is undertaken. A professional survey organisation will normally be contracted to administer both the pilot surveys and the main survey. Such companies have experience in market research and related fields, and use interview staff who are familiar with the application of questionnaire methods and with handling respondents in survey situations.

11.08 Whether survey responses are collected directly on laptop PCs or on paper forms, the results need to be collated and stored securely for possible future reference. The data need to be reliably transferred (with cross checking for transcription errors) to a secure medium which can then be made available for data analysis, which is the subject of Chapters 12 and 13.

Table 11.1 **Types of bias in stated preference analysis**

Type of bias	**Nature of bias**	**Effect on WTP (*t*WTP = 'true' WTP)**	**Solutions**
Hypothetical	Scenario is not consistent with reality	WTP $\neq$ *t*WTP	Design a plausible scenario (see Chapters 9 and 10)
Strategic — classic free rider	If respondent believes payment of their WTP will be collected from them:	WTP $<$ *t*WTP	Remove outliers. Seek motivations for WTP. Use referendum (yes/no) format
Strategic — free rider	If respondent believes payment will only be collected from others	WTP $>$ *t*WTP	As above. Strategic bias not thought to be serious in practice
Starting point bias	WTP anchored on initial stated value	WTP = initial value $\neq$ *t*WTP	Use open ended or payment card techniques
Framing effects	WTP depends on how the question is framed (e.g. whether 'the glass is seen as half full or half empty')	WTP varies with frame when it should be the same for the same good	Possible presentation of questions in a 'neutral frame'
Payment vehicle bias	WTP depends on how the good is to be financed	WTP *for the good alone* should be invariant with payment vehicle	May not be a problem if the good is redefined to include the method of payment. Otherwise, payment vehicle should be as close to the one that would be used in reality
Embedding/ scope insensitivity	WTP does not vary with quantity of good offered, or when first quantity of good is included in a second quantity which has another good as well	WTP may reflect *warm glow*, i.e. satisfaction of paying for the good, not the WTP for the good itself.	Although there are disputes about whether warm glow is a problem, and about the validity of studies finding warm glow effects, the wording and followup should be designed to ensure that the full WTP is captured
		Could reflect satiation: i.e. having enough of the good	Satiation is legitimate so not a problem
Sensitivity to sequencing	WTP varies with where in a sequence of goods a particular good appears	WTP $\neq$ *t*WTP?	Sequencing effects may be rational. One would expect lower WTP the later a good appears in the sequence since disposable income will be less given that first sequenced goods have been purchased
Yea saying	Respondent tries to please the interviewer	WTP $>$ *t*WTP	If detected, calibrate responses when analysed by scaling down
Nay saying	Respondent concerned to counter the interviewer	WTP $<$ *t*WTP	As above, but scale up
Protest responses	Refusal to answer, or may give ludicrously high WTP or untrue zero WTP ('protest zeros')	No WTP WTP $>$ *t*WTP WTP $<$ *t*WTP	Ask for reasons for refusal to state WTP (follow up questions). Avoid open ended formats. Omit outliers especially where WTP exceeds income. Examine reasons for zero WTP: many zeros will be legitimate zero bids. Omit 'true' protests (see Chapters 12 &13)
Preference imprecision	Respondent unable to cite precise WTP	There is no *t*WTP but a range	Record the range (e.g. using a payment ladder). Ranges are not signs of bias but a fact of life
		Could reflect underlying randomness of preferences	Underlying theory is breached, so any resulting WTP could be a construct of the questionnaire
Reference dependency	WTP varies with the reference point as perceived by the respondent	Most likely difference between questions framed as WTA rather than WTP: WTA $>$ WTP	Test for reference dependency by asking both WTP and WTA questions. If there is aversion to money, use CM. WTA more relevant when property rights vested in losers, and WTP when vested in gainers
Information effects	WTP varies with information provided	*t*WTP corresponds to 'right' amount of information?	Not a bias: one would expect WTP to vary with information, as with any market good. All prices are conditional on context

PART 4

Analysing stated preference data

CHAPTER 12
Analysing the data: contingent valuation

Summary

Data analysis requires a number of preliminary steps prior to the main tests. These summarise the data and particularly the WTP/WTA responses to give an accessible overview, and identify and eliminate responses which are not valid. Weighting of the results may be required if sampling is not random, and upper and lower bounds for the WTP/WTA values need to be identified. The two main aims of data analysis are then explained: to estimate the mean and median WTP/WTA of the sample with associated confidence intervals, and to estimate a bid function to test that responses have a distinguishable structure and conform to prior expectations and economic theory. Finally, the case for estimating a benefit transfer equation is explained.

Analysing the data generated by SP surveys is a complex statistical task, and will require professional advice from econometricians. This guide does not provide full details of such procedures, although more information can be found in the Manual.[9] Furthermore, as in any context, using a peer review process for quality assurance may well increase the degree of confidence in the final results.

12.01 The analysis stage of a contingent valuation (CV) study is the point at which data collected in the survey are transformed into a useable output.

12.02 The key importance of high quality questionnaire design was discussed in Chapters 9 to 11, but it is also essential that equivalent attention be paid to the quality of the data analysis. Indeed, one should have as little confidence in the results of a high quality survey that has been poorly analysed, as one should have in the results of a poor survey that has been competently analysed.

12.03 Before the analysis of data can begin, four tasks must already have been completed, as we have seen:

- the *population* whose aggregate WTP/WTA is required has been defined (Chapter 8);
- a *sample* from that population has been selected according to a sampling rule (Chapter 8);
- a *CV questionnaire* eliciting WTP/WTA for a change in provision of a good has been designed (Chapters 9 and 10); and
- after the questionnaire has been tested and then administered to the sample, the resulting survey data have been tabulated in electronic format and cross-checked for input errors (Chapter 11).

9 I. Bateman, 2002, *op. cit.*

12.04 Analysis of data collected in a CV study should achieve the following objectives:

- Estimate the mean and median WTP/WTA of respondents in the sample;
- Test for structure in the WTP/WTA responses to the CV survey. That is to ascertain whether the differences in responses can be explained through differences in the characteristics of the respondents; and
- Provide a transfer equation to assist in subsequent benefit transfer exercises.

12.05 Particular studies may have additional objectives, such as investigating the impact on WTP/WTA of changes in the scenario — i.e. the hypothetical policy or project — which is presented to respondents. However, this chapter deals solely with the three objectives listed above, after first considering the preliminary steps of organising the survey data.

Preliminary steps

Summarising the data

12.06 Analysts should begin their analysis by summarising the data collected in the CV survey. The best way to summarise the responses to any particular question will depend on the nature of the question. In general, however, analysts should attempt to provide details of the key characteristics of the responses to each question in a way that is readily accessible to those commissioning and reviewing the study. This will usually entail the use of a variety of numerical and graphical devices.

12.07 Special attention should be paid to summarising responses to the WTP questions. Naturally, this summary will depend on the type of elicitation format that has been used in the survey. For the purposes of analysis, the different elicitation formats result in data that can be grouped into three basic categories. These are summarised in Table 12.1.

Table 12.1 **Types of WTP data collected in CV surveys**

Data Type	Elicitation Question type	Description
Continuous	Open-ended Bidding game	Each respondent states one amount that corresponds to their maximum WTP
Binary	Single-bounded discrete choice	Each respondent states whether their maximum WTP is above or below a given amount
Interval	Double-bounded discrete choice Multiple-bounded discrete choice Payment card (and similar approaches)	Each respondent identifies two amounts that bound their maximum WTP (i.e. one amount greater than and one less than their maximum WTP)

Identifying non-valid responses

12.08 The first step in the analysis of CV data is to establish which responses can be treated as valid reflections of respondents' true WTP.[10] In general, responses which correspond to any of the following categories should be classified as non-valid:

- those refusing to answer the valuation questions;
- those which do not provide their genuine WTP but respond with a zero value instead, and
- those which do not provide their genuine WTP but respond with an unrealistically high value instead.

See Table 9.7 for some criteria for judging response validity.

12.09 Once the study has reached the analysis stage, it is clearly too late to avoid households offering non-valid responses. However, it is the analyst's responsibility to use the available data to decide which WTP values should be considered genuine reflections of a household's valuation of the good being offered in the contingent market and which should be considered non-valid responses.

12.10 The questionnaire should provide at least five sources of information to aid this decision:

- follow-up questions asking why no response was offered to the WTP questions;
- follow-up questions asking why a zero WTP was offered. Zero responses may be genuine, but alternatively could reflect strategic behaviour of low WTP respondents;
- debriefing question on the confidence the interviewer had regarding the interviewees' answers;
- scope for comparing very high bids with respondent's uncommitted income (WTP cannot exceed ability to pay), and
- debriefing of interviewers who may notice reasons for this behaviour.

12.11 The basis for decisions concerning the identification of non-valid responses should be clearly detailed. Once all responses have been scrutinised for validity, the analyst should tabulate how many responses have been categorised as non-valid, and for what reason.

12.12 Non-valid responses often reflect respondents' objections to aspects of the contingent valuation scenario (e.g. being asked to pay for the provision of something that they believe they should have free by right). Such objections should be identified in pre-testing, and where possible the scenario should be modified so as to minimise the potential for non-response. Surveys that report high levels of non-valid responses should be treated with some scepticism.

10 In this context, the word 'valid' is used to denote WTP responses that conform to the economic principles underlying CV research and can be included, within that framework, in the analysis of the benefits from a policy or project.

12.13 Ultimately, the WTP of those returning non-valid responses is not known. The only course of action open to the analyst is to delete these respondents from the sample. However, in deleting observations analysts must ensure that the reduced sample does not differ significantly in its characteristics from the unadjusted sample. If the characteristics of the sample have changed, weighting procedures should be used to compensate for the unrepresentativeness of the reduced sample.

Choosing analytical weights

12.14 The data collected in a CV survey contains information on the WTP of a sample of respondents. This sample is merely a subset of a larger target population. As described in Chapter 8, it is extremely important to ensure that the sample is representative of the population in general. Indeed, the results of the survey can only be extrapolated accurately and reliably to the entire population if the sample is an unbiased subset of that population.

12.15 Ideally, the data would result from a simple random sample in which each member of the population had an equal probability of being in the sample. Such a sample can be considered to be a representative sample.

12.16 If the sample is skewed such that certain members of the population are more likely to be in the sample than others, then the sample will be biased and not representative. Sample bias occurs when:

- observations that have a low probability of being in the sample are attributed an above average weight and hence have a greater than average influence on the analysis, and
- observations that have a high probability of being in the sample are attributed a below average weight and hence have a less than average influence on the analysis.

In either of these cases, **analytical weights** that correct for this problem must be defined and applied to the sample data.

12.17 There are four possible cases, each of which requires a different approach to the use of analytical weights:[11]

- If a **simple random sample** is used, then each member of the population had an equal probability of being in the sample. Hence each observation should be allotted the same importance in the analysis and, as such, there will be no need to define weights.
- If the sample has been selected using a **probabilistic process** (e.g. by stratified or clustered sampling, see Chapter 8) then weights will be relatively easy to define. Each observation's weight can be deduced from the probability of that respondent being selected for the sample and this probability should be known from the sampling design.
- If sampling has been **non-probabilistic** then some members of the population may have a strong chance of being chosen, others may have no chance at all, and there is no way to tell the specific probability in either case. Though such samples are difficult to

11 These are explained in Annex 5.1 of I. Bateman (2002) *op. cit.*

interpret, weights should be defined which attempt to improve the representativeness of the sample.

For example, imagine that following the removal of **non-valid responses** an examination of the characteristics of the sample shows that the proportion of households in the sample from low income groups is significantly lower than that in the population. The solution is to calculate analytical weights for households in each income group by dividing the proportion of the *population* falling into that income group by the proportion of the *sample* falling into that income group.

Much more complex weighting procedures exist. For example categories may be defined according to more than one characteristic (e.g. income and age), in which case the same procedure applies: define the strata, find the population proportion in each stratum and divide these by the proportion of the sample in the equivalent stratum.

- If it is **impossible to define sampling weights,** then analysts must proceed without them and rely on their own subjective evaluation of the survey results.

Determining admissible bounds for WTP/WTA

12.18 Great care is needed in the design of CV questionnaires to ensure that households are responding to WTP/WTA questions in a way that is consistent with welfare economic theory. It is vital that the analysis of data from CV surveys maintains this consistency by ensuring that appropriate bounds are identified for responses.

12.19 To establish how economic theory can guide the analysis of the data, it is necessary to answer a number of questions about nature of the valuation exercise:

- Does the scenario in the CV questionnaire present the respondent with the prospect of an increase or a decrease in the provision of the non-market good?
- Have the valuation questions elicited a WTP or WTA value?
- For WTP data, how frequently would the household make their hypothetical payments?
- Does it make sense for households to have negative WTP/WTA values?

12.20 In answering these questions, the analyst is able to determine the upper and lower bounds that might be placed on the responses. For the upper bound:

- If the data report households' WTA to forgo an increase or to suffer a decrease in provision of the non-market good, then there is no theoretical limit to the amount that households might wish to be compensated. We cannot put a theoretical upper bound on admissible WTA values (but see the section above in this chapter on Identifying Non-Valid Responses).
- If the data report households' WTP to achieve a gain or to avoid a loss in provision of the non-market good, then we know that they can only pay as much as they can afford. The analyst should attempt to identify the relevant measure of income to use as an upper bound. If data are a monthly or annual payment then the relevant upper bound

for WTP data will be the household's discretionary income for that period. Discretionary income is that part of income that is left over once committed expenditures on housing, food, clothing, transportation and other market and non-market goods have been taken into account. Unfortunately, it is notoriously difficult to obtain accurate estimates of discretionary income from survey respondents. Analysts may have to resort to approximations based on reported gross household income (e.g. taking a percentage of after-tax income).

12.21 For the lower bound:

- If the item being valued is a public good — i.e. one where the benefits cannot be restricted to those who pay — then it frequently makes little sense to talk about negative WTP/WTA since the good can simply be ignored if it does not provide utility to the respondent. In many cases, we can impose the restriction that WTP/WTA values must be non-negative.

- In other cases, the issue may not be so clear. In general, however, the CV survey should be designed to identify respondents that like, dislike or are indifferent to the proposed change in provision of the non-market good. In so doing, the valuation question can be framed appropriately. For example, respondents experiencing a welfare loss can be asked to reveal the minimum they would be willing to accept in compensation for the change, whilst those experiencing a welfare gain can be asked the maximum they would be willing to pay for the change. Again, it makes sense to analyse each sub-group separately under the assumption that WTP/WTA values must be non-negative.

12.22 Lower and upper bounds should be defined for each observation in the data set, where this is practicable.

Estimation of mean and median WTP

12.23 The key objective in analysing CV data is to obtain estimates of two summary statistics: the *sample mean WTP* and the *sample median WTP*:

- The mean WTP is what in common usage would be termed the average WTP of the sample, and

- the median WTP is the value of WTP that divides the sample exactly in half (that is, the value of WTP at which exactly 50% of the sample have a lower WTP and 50% have a higher WTP).

12.24 For policy makers the two measures for summarising the sample's WTP have quite different interpretations. The mean will be relevant if the context of the valuation exercise is cost-benefit analysis. If mean WTP for benefits outweighs costs, this suggests that the project should proceed. The median value is relevant in the context of public choice (i.e. the economic analysis of political decisions) since it corresponds to that amount which would just receive a majority approval.

12.25 Some analysts have argued that median WTP is a more robust measure of central tendency since its value is not so greatly influenced by outliers. However, both mean and median

values of WTP should always be reported, since neither measure is innately superior. Mean and median WTP may be estimated by using either non-parametric or parametric models. The distinction is less important than the fact that they both provide complementary views of how the data fits the model.

Mean and median WTP from non-parametric models

12.26 Mean and median WTP should first be estimated using **non-parametric estimation techniques**. This is because non-parametric estimation is a set of techniques that can be used to estimate mean and median WTP while making almost no assumptions about the nature of the WTP data. In some cases these estimates can be obtained through the simplest of numerical calculations.

12.27 The particular techniques which are appropriate to use depend on the type of data which have been collected in the CV survey (see table 12.1) as follows:

- *continuous data*: the Kaplan-Meier estimator;
- *binary data*: the Pooled Adjacent Violators algorithm; and
- *overlapping interval data*: Turnbull's Self-Consistency algorithm.

Professional advice may be needed here. The techniques are explained in detail in Annex 5.2 of the Manual.

12.28 The importance of non-parametric estimation cannot be overstated. The mean and median WTP values derived from non-parametric estimation can be taken as lower-bounds for these statistics. In other words, they give the minimum value for the mean or median that is consistent with the sample data.

Mean and median WTP from parametric models

12.29 Analysts should also provide 'best estimates' of mean and median WTP, and one way of achieving this is to turn to the techniques of parametric estimation. With parametric estimation, the analyst assumes that WTP is distributed in the population according to some probability distribution. The objective of the estimation exercise is to find the parameters of the assumed probability distribution function.

12.30 In the first instance the analyst must choose a probability distribution with which to model WTP. Once the probability distribution function has been selected, the analyst is in a position to estimate the parameters of the model. Once a parametric model of WTP has been estimated, mean and median WTP can be calculated. For many distributional assumptions, they can be calculated using straightforward formulae. Median WTP can always be calculated with relative ease: the analyst simply has to solve for the value of WTP at which the cumulative density function of the assumed distribution evaluates to 0.5. For the calculation of mean WTP for complex distributional assumptions, there may be no

convenient formula. In these circumstances it may be necessary to resort to methods of numerical integration in order to derive mean WTP from the estimated distribution function.

Calculation of confidence intervals for mean and median WTP

12.31 In calculating mean and median WTP values, the results refer just to one sample of households. They are simply *estimates* of the mean and median of the entire population based on the information provided by the sample. A separate sample of households would return different WTP values and we would end up with different estimates of the population's mean and median WTP. Thus, analysts should also provide an indication of the accuracy of these estimates. This requires the construction of a 95% confidence interval. In general, there are two approaches to determining confidence intervals: an **analytical approach** based on statistical theory, and a **numerical approach** based on extensive computing.

12.32 In general, however, analytical formulae for calculating confidence intervals exist only for relatively simple models. Instead, analysts frequently resort to numerical techniques, and professional advice on this should be sought from an econometrician. The two techniques that have enjoyed most widespread application are the **Krinsky-Robb** method and **bootstrapping**. These can be summarised as follows:

- the Krinsky–Robb method can only be used with parametric models and relies on certain assumptions concerning the distribution of the parameters of the estimated model, and

- bootstrapping is a very robust technique that can be used to construct confidence intervals for mean and median WTP using any type of data, i.e. continuous, binary or interval or results from any estimation method, i.e. non-parametric or parametric.

12.33 Of the two methods, bootstrapping is to be preferred. Though computationally burdensome, the advantages are that the confidence intervals constructed make almost no assumptions concerning the nature of the data and the technique is applicable in practically all situations.

Testing for structure in WTP responses

12.34 The second objective of the analysis of CV data (after estimating sample mean and median WTP) is to establish the validity of the WTP responses. Analysts will thus need to test whether the WTP values provided by respondents follow distinguishable patterns and whether these patterns conform with prior expectations and economic theory.

12.35 In general, analysts will have a number of variables that they conjecture will influence the WTP of a respondent. Typically such variables will include:

- the respondent's income and other socio-economic characteristics;
- details of their attitudes towards the programme offered in the CV scenario;
- information on their current knowledge of the good to be provided, and
- for goods with a spatial dimension, their proximity to the site of provision.

12.36 The objective is thus to see how well these variables 'explain' households' WTP. The approach is to specify, estimate and then interpret a **bid function**. The bid function is a mathematical equation that, once estimated, describes how each variable influences a respondent's WTP.

Specification of the bid function

12.37 In general, the bid functions used in CV analysis are relatively simple. Typically, one parameter is associated with each variable such that, once estimated, the sign of the parameter indicates whether increasing values of that variable have a positive or negative influence on a household's WTP, and the significance of the estimated parameter determines whether it is possible to attribute any statistical significance to this influence.

12.38 Two alternative paradigms can be followed in specifying the form of the bid function:

- **Utility difference models.** The bid function is modelled as the monetised value of the *change in utility* experienced by a respondent following a change in provision of a non-market good. A respondent's income is assumed to influence the monetary value of each unit of utility change. The other variables are assumed to influence the size of the utility change experienced by the respondent, and
- **Bid function models.** The bid function is specified directly. Income and the other variables are simply assumed to influence the size of respondent WTP.

12.39 There is no clear consensus in the literature regarding which paradigm should be adopted in determining the bid function. The utility difference model is consistent with economic theory, but produces more complex forms of the bid function. The direct bid function models produce more tractable bid functions but have less affinity with underlying theory. In general, it is good practice for analysts to choose one paradigm and include in their report a brief description of the economic theory underlying the chosen approach.

Estimation of the bid function

12.40 The bid function is estimated using the fully parametric techniques summarised above. Once again the analyst must choose a parametric distribution of WTP that does not contravene economic theory. As with the choice of probability distribution function for models used to derive estimates of mean and median WTP, this will frequently require that the distribution only supports **non-negative values lower than the households' discretionary income**.

12.41 Importantly, analysts should be aware that the 'best fitting' model when covariates are included does ***not*** have to make the same distributional assumptions as that used in a parametric model used to evaluate mean and median WTP.

12.42 Choosing a 'best fitting' model is not an exact science. Analysts' choice between different specifications of the bid function and distributional assumption will be guided by the signs and significance of the various parameters and the overall fit of the model. Analysts should avoid 'mining' the data. Frequently, the factors determining respondent's WTP are extremely complex, and it is unlikely that these will be adequately captured by a simple bid function.

Interpretation of the bid function parameters and overall explanatory power of the model

12.43 Unlike other fields of econometric analysis, CV analysts do not tend to be particularly concerned with the actual values of the estimated parameters. Since the objective is to establish that the respondent's WTP values are not purely random, CV analysts tend to focus more on the sign and significance of the estimated parameters and explanatory power of the whole model.

12.44 The analyst should check that estimated parameters have signs that conform to prior expectation. Thus, variables such as income that are expected to increase WTP should be **positively signed** whilst variables such as distance to the site of provision which are expected to reduce WTP should be **negatively signed**.

12.45 The statistical significance of these parameters can be tested using a simple ***t*-test**. The statistical significance of each parameter should be reported. A measure of the explanatory power of the entire model is provided by the **pseudo R^2 statistic**. This statistic takes values between 0 and 1, where a value of zero suggests that the included covariates do nothing to explain the distribution of WTP in the sample.

12.46 In general, the larger the value of the pseudo R^2 statistic the greater the explanatory power of the model. Unfortunately, there is no commonly accepted threshold value for the pseudo R^2 statistic that denotes a satisfactory or well-specified model. However, analysts should be concerned if the inclusion of covariates provides very little explanatory power. Certainly one would be concerned if the pseudo R^2 statistic were less than 0.1 and one might then draw the conclusion that the WTP values returned from the CV survey show very little in the way of distinguishable patterns.

The benefit transfer equation

12.47 One objective of the analysis of CV data should be to provide details that can be used in the transfer of benefit estimates from the current study to other similar contexts.

12.48 It may prove adequate simply to transfer the average WTP value, though this value is only strictly applicable to the population from which the sample for the study was drawn. So to improve the process of benefits transfer, analysts may wish to estimate a **transfer equation**, which measures WTP as a function of respondents' characteristics. Given that the characteristics of the population to which the estimate will be transferred are likely to differ

from those of the study population, it is hoped that benefits estimates can be improved by using the transfer equation to modify the estimate of average WTP to account for these differences. Such a model will contain only a limited number of covariates, reflecting characteristics that can easily be gathered for the transfer population. Typically, this amounts to basic socioeconomic details, such as respondents' income, age and sex, which can provide a basis for comparison and adjustment at the destination site.

12.49 For those goods that have a spatial dimension, it is essential to include a variable that measures households' distance from the site of provision. The parameter estimated on such a variable is crucial in establishing the rate of distance decay of WTP and hence in defining the boundaries of the population that may have a positive WTP in a transfer exercise.

12.50 It should be standard practice to present the results of such a benefit transfer model in a report on the analysis of data from any SP study, as this may help future researchers to make use of the information obtained.

CHAPTER 13
Analysing the data: choice modelling

Summary

The issues in analysing choice modelling data are similar to those dealt with for CV (in Chapter 12), and require establishing welfare estimates for the median and mean WTP/WTA of the sample, along with relevant confidence intervals. This is done for CM, as with CV, by selecting and then estimating a model. Focussing on choice experiments (CE) as a sub-category of CM, this chapter explains random utility theory and how econometric models based on it are used for data analysis.

CE techniques in particular regard any good as a bundle of characteristics, and they differ from those used for CV by using a random utility function to depict the mathematics of consumer choice. A range of random utility models can be used, based on decisions about the appropriate functional form of the indirect utility function and the probability distribution of the random elements. Estimation of the model from CE data will then enable the change in utility from the hypothetical change in attributes of a good to be expressed as a money value. Testing the responses for structure or validity follows the same procedures as for CV, as outlined in Chapter 12.

The warning from the start of Chapter 12 also applies here: analysing the data generated by SP surveys is a complex statistical task, and will require professional advice from econometricians. This guide does not provide full details of such procedures, although more information can be found in the Manual. Furthermore, as in any context, using a peer review process for quality assurance may well increase the extent of confidence about the final results.

13.01 The term choice modelling (CM) encompasses a range of approaches including choice experiments, contingent ranking, contingent rating, and paired comparisons, as described in Chapter 10. These approaches are based on the idea that any good can be described in terms of its attributes or characteristics and the levels that these take. For the purposes of choice modelling, one of the attributes will be the cost of providing that good.

13.02 The various methods present respondents with descriptions of one or a number of goods, differing with respect to the levels of attributes and the costs of provision. Based on the descriptions of the attribute levels and costs, respondents are asked to choose between the different goods on offer and the status quo. The way in which the respondent is asked to make this choice defines the CM technique. For example, a contingent ranking exercise would ask respondents to rank the different options in order of preference.

13.03 As described in Chapter 10, choice experiments, in which respondents are asked to select their most preferred option from the different options on offer, are fully consistent with economic theory. This chapter therefore focuses solely on the analysis of CE.

13.04 The organisation of data is an extremely important task, especially in CM contexts where data do not take on continuous form. Each record must contain details of the levels of attributes of each of the options presented to a respondent as well as a dependent variable that indicates which of the options was selected. The particular form of the data will depend on the econometric package being used to estimate the model. Quality control

procedures would normally be used to screen for incomplete data records, outliers, and obviously incorrect entries.

13.05 To analyse CE data and obtain mean and median WTPs with confidence intervals, we require an econometric model that can be used to describe discrete choice behaviour. This is provided by random utility theory, which is derived from the economic assumption that a rational individual will select the option from a set of options that provides them with the greatest expected utility.

13.06 Further, CE makes the assumption that the utility an individual would derive from any one option will depend upon the level of attributes provided by that option and the cost (to the individual) of providing it. The level of utility associated with an option will vary between different individuals.

13.07 The first step in formulating the econometric model is to specify an **indirect utility function** that relates attribute levels, costs and individual characteristics to the level of utility enjoyed. Analysis of CE data seeks to determine the parameters of this function based on observations of how individuals choose between options.

13.08 The indirect utility function used by the analyst is merely an approximation to the individual's actual indirect utility function. As a result the utility predicted by this approximation will almost certainly not be 100% correct. To complete the econometric model, we must add a **random element** to the analyst's indirect utility function. This error component picks up the difference between the true utility of an option and that which is modelled. The inclusion of this random element explains the derivation of the title **random utility model (RUM)**. Since there is an error part of the utility function, the analysis becomes one of **probabilistic choice**. The probability that any particular respondent prefers an option to any alternative option can be expressed as the probability that the utility associated with that option — according to the model — exceeds that associated with all other options.

13.09 In formulating a random utility model, the researcher makes two decisions:

- The **functional form of the indirect utility function:** In practice researchers usually opt for extremely simple specifications of the indirect utility function. Indeed, in the vast majority of applications the utility of an option is modelled as a simple linear combination of costs and attributes.

- The **probability distribution of the random elements:** A typical distribution in this case is the Gumbel or extreme value distribution, which is similar to the normal distribution but has more tractable mathematics associated with it. The resulting probability that a particular choice is made can be expressed in terms of the logistic distribution. When the choice is between *two* options, the model is known as a binary logit model. When the dependent variable takes *three or more* values, as is usually the case in CE, the model is a conditional (or multinomial) logit model. If a generalised extreme value distribution is chosen instead of a Gumbel distribution, the resulting model is the nested multinomial logit model. If, instead, a normal distribution is chosen, then the resulting model is a binary probit model in the case of two options, or a multinomial probit model for several options.

13.10 In terms of mathematical tractability the simplest model is the **conditional logit model** resulting from the assumption of Gumbel distributed error terms. Unfortunately this model is built upon the assumption of the **'independence of irrelevant alternatives' (IIA)**. IIA

stipulates that the ratio of the probabilities of choosing any two options will be unaffected by the attributes or availability of other options. In reality we would expect IIA to be quite easily violated since the introduction of a third option will nearly always alter the relative probabilities of choosing the first two options.

13.11 If the analyst opts to use the conditional logit model then a formal test for the IIA should be carried out. If this fails, then the analyst should consider choosing an alternative model. The **nested logit model** somewhat relaxes the IIA assumption by grouping more similar options together. The **multinomial probit model** is the most general model and makes no IIA assumption.

13.12 Probabilistic models such as the ***binary*** or ***multinomial logit model*** or the ***binary probit model***, can be estimated using **maximum likelihood procedures**. These are available as standard routines in nearly all econometric software packages.

13.13 The **nested multinomial logit model** is also estimated using maximum likelihood procedures. Unfortunately few econometric packages provide this as a standard routine (though LIMDEP has an add-in package that includes a nested multinomial logit estimator). However, it is possible to use a stepped procedure (the number of steps depending on the number of levels of nesting) using a multinomial logit estimator which approximates the full nested model.

13.14 The **multinomial probit model**, being the most complex of the RUMs presented, requires the use of **simulated maximum likelihood procedures**. Such procedures are not generally available as standard routines (though once again an add-in package is available for LIMDEP).

13.15 Whatever Random Utility Model has been chosen is estimated by econometric analysis to provide the mean and median WTP/WTA information, with confidence intervals. Data from a CE allow for the estimation of the indirect utility function (which is thus a simplified approximation to the real thing). Welfare measurement, therefore, progresses by the two steps of:

- assessing how respondents' utility would change if the attributes of a non-marketed good were changed from their current level (the status quo) to some different level (the levels set by the policy option under review); and then,

- expressing this utility change in money terms.

13.16 Notice that by varying the levels of attributes in the policy option under review it is possible to estimate the welfare benefits of any number of different programmes. This is a strength of CM when compared to CV.

13.17 When a simple linear functional form is used for the indirect utility function these formulae are very straightforward. Also, it is simple to calculate **marginal WTP for any one of the option attributes** according to:

$$\text{MWTP} = -\frac{b_k}{\beta}$$

where β is the coefficient on cost and, b_k is the coefficient on attribute k.

13.18 Notice the intuitive interpretation of this ratio: b_k is the utility from an extra unit of attribute k; β is the value in money terms of one more unit of utility. Dividing one by the other gives the monetary value of the utility coming from an extra unit of attribute k. This ratio is also referred to as the **implicit price** of the attribute.

13.19 As with CV data, analysts should always report 95% confidence intervals on these estimates of welfare changes. For simple specifications of the indirect utility function it is possible to employ analytical formulae to calculate confidence intervals.

CHAPTER 14
Reliability and validity: contingent valuation and choice modelling

Summary

This chapter discusses the issues involved in finding out whether a particular contingent valuation or choice modelling questionnaire has provided an unbiased and transparent vehicle to give respondents the best chance to deliberate about their preferences and to get as close as possible to the values they would affirm in the light of experience. Tests of validity are applied to decide whether the 'biases' discussed in Chapter 11 exist. Validity refers to the degree to which a study succeeds in measuring the intended values by overcoming potential biases and the hypothetical nature of the exercise. In general, tests of validity and reliability will be the same for choice modelling as they are for CV experiments, although CM allows for further tests of validity.

Several types of validity are identified and explained: content validity, and two kinds of construct validity, namely convergent validity and expectations based validity. A number of survey design issues bearing on validity are also discussed.

14.01 The objective of a stated preference (SP) survey is to elicit respondents' WTP or WTA for the change in provision of a non-market good described to them in the scenario. It is anticipated that respondents will state the amount that they genuinely believe they would be willing to pay or willing to accept in compensation if the change in provision actually occurred. This value is called the respondent's formulated value. However, if respondents perceive some strategic advantage in mis-reporting their values then their stated value may not equal their formulated value. Clearly, a prime objective of questionnaire design will be to ensure that the questions encourage truth telling such that it is in the respondent's interest to ensure that stated value equals formulated value.

14.02 We hope that the amount the respondent genuinely believes they would be willing to pay or willing to accept when answering an SP question is the same amount as they would actually pay if they were given the opportunity in a real world exchange. In other words, we hope that their formulated value is the same as their actual value. Again, there are reasons why we might expect these values to differ. Most importantly, differences in the two values may arise because respondents expend far greater cognitive effort upon determining actual as opposed to formulated values.

14.03 The task of the questionnaire is to provide an unbiased and transparent vehicle which gives respondents the best possible chance to deliberate about their preferences and approach as closely as possible to the values that they would affirm in the light of experience. The criteria upon which success is judged can be divided into tests of reliability and tests of validity. **Reliability** refers to the degree of replicability of a measurement. That is, can a survey instrument be relied upon to provide the same values if we were to administer it repeatedly under controlled conditions. **Validity** refers to the degree to which a study succeeds in measuring the intended quantity. That is, to what extent has the survey

instrument overcome issues of bias and the hypothetical nature of the exercise to arrive at respondents' actual values.

14.04 Reliability exercises typically entail the repetition of studies at different points in time and so are not considered to be a reasonable requirement for each individual study. The literature to date is on the whole supportive of the temporal reliability of SP results.

14.05 The central problem in assessing the validity of WTP/WTA values obtained from any stated preference study is the absence of a definitive yardstick against which to compare those measures. This is not a generic problem of all survey research (e.g. election opinion polls can be compared against the results from the subsequent elections they set out to predict). However, it is generally a problem for non-market goods since, with very few exceptions, actual values are unobservable.

14.06 Testing the validity of a particular SP valuation survey therefore requires more indirect means (see Figure 14.1). In general, validity is assessed by two criteria:

- Whether the SP survey asked the right questions in a clear, understandable and appropriate manner, which is termed **content (or face) validity;** and,

- Whether the values produced by the SP study follow the patterns we expect (e.g. WTP increases with household income, if everything else remains the same) and are in accordance with values derived from other studies, which is termed **construct validity**.

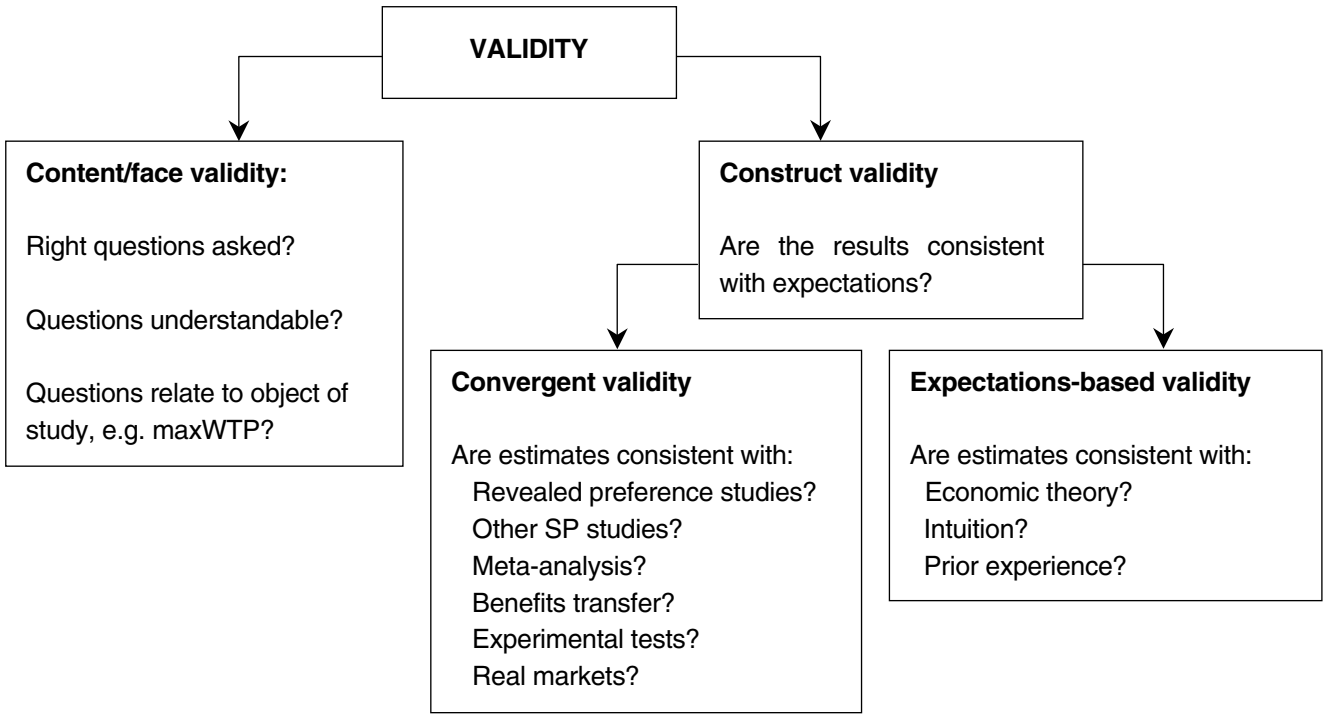

Figure 14.1. Types of validity tests

Content validity

14.07 A study with high content validity would be one in which the survey descriptions and questions are clear, reasonable and unbiased, such that respondents are put in a frame of mind that motivates them to answer seriously, thoughtfully, and, of course, truthfully. These issues need to be considered at the questionnaire design and testing stage (see Chapter 11), but it is equally important to address the same issues when the survey data is analysed.

14.08 Assessing the content validity of a survey is essentially a subjective expert appraisal task. Content validity judgements encompass the entirety of the study, from considering how reasonable the aims of the study are, to the clarity, interpretability and plausibility of the questions posed, and to the way in which the survey was conducted. The guiding principle should be whether the diversity of components which constitute an SP study (the sampling frame, administration of the survey, the information and materials provided to respondents, the description of the good and associated provision change scenario, etc.) are conducive and sufficient to induce respondents to reveal valid stated values. The following questions provide a checklist of general components of an SP study that should be examined to assess content validity:

Issues of scenario design:

- Is the good offered clearly specified to and understood by respondents?
- Is the information provided adequate and reasonable to describe the provision change and payment scenario?
- Is the trade-off between money and the good plausible?
- Are substitutes and the consequences of non-payment adequately described?

Elicitation issues:

- Is the chosen measure of wellbeing appropriate (WTP or WTA)?
- Is the chosen elicitation format appropriate?

Institutional context:

- Are the methods of provision and allied institutional arrangements plausible?
- Are respondents likely to have an expectation of having to pay for the good if it is provided?
- Are respondents likely to feel that they are providing an input to the decision-making process?

Sampling:

- Has the correct population been identified and adequately sampled?

Survey format:

- Is the choice of survey mode appropriate?
- Has the survey administration and data preparation been conducted to a sufficiently high standard?
- Does the questionnaire design collect adequate data concerning variables that are likely to explain WTP, so as to permit construct validity testing (including the elicitation of attitude and response reason data)?

14.09 Ideally an SP questionnaire should successfully pass all the tests of content validity summarised above. Indeed, the results of surveys that fail to achieve these standards should be treated with scepticism. The following list presents some indicators of SP studies that have low content validity:

- inadequate sample size or poor coverage of the relevant population;
- non-stratified or biased sampling methods (where the representativeness of the sample is an issue);
- high survey or individual question (or item) non-response rates;
- large numbers of 'protest' bids;
- prevalence of free-riding behaviour;
- high numbers of infeasibly large bids;
- inadequate responsiveness to the scope of the good in question, i.e. WTP does not vary with quantity of the good;
- the valuation scenario and corresponding valuation task is poorly understood and/or has low credibility;
- the description of the change in the provision of the good is poorly understood and/or has low credibility;
- the relevant authorities are not trusted or considered to be of low competence or efficiency;
- low explanation of responses in terms of theoretical or other expectations; and
- survey or post-survey respondents provide answers which indicate that strategic behaviour may have affected responses (e.g. respondents under- or over-state WTP).

14.10 One indicator that deserves further discussion is that of protest bids. Protests may show up as a refusal to answer the valuation questions in an SP survey, i.e. not 'playing the game'. They may also involve respondents who say WTP is zero when in fact it is greater than zero (protest zeros), and those who give unrealistically high bids. The use of professional interviewers can minimise refusals. Ways to minimise the number of protest bids include:

- avoiding open-ended elicitation formats which tend to be associated with high levels of protest zeros;

- follow-up questions to ask why a zero WTP was offered. Zero responses may be genuine, but could reflect strategic behaviour of low WTP respondents. Scenarios may be poorly formulated and this needs to be checked in the follow up questions;
- comparing very high bids with the respondent's uncommitted income (WTP cannot exceed ability to pay);
- debriefing interviewers who may notice reasons for the protest bids;
- inferring the WTP of any respondent who does not answer the valuation question but answers other questions. Inferred values could be taken from the WTP statements of respondents who have similar characteristics (income etc) to those who protest and
- deleting non-respondents and adjusting the sample to reflect any change in representativeness.

Convergent validity

14.11 Convergent validity assessments typically compare measures obtained from SP studies with:

- results from other methods (e.g. travel cost or hedonic pricing);
- results from other SP studies (e.g. via meta-analysis and benefit transfer assessments), and
- actual or surrogate (proxy) markets.

Where possible such comparisons should be made and reported.

14.12 However, caution should be exercised in the interpretation of convergent validity tests. In the majority of cases, neither the value derived from the SP study nor the value against which this is being compared can automatically claim superiority in terms of being a naturally closer approximation of the 'true' value. Just because two studies deliver similar or logically related measures does not mean that those measures are valid; instead they may be equally invalid. Nevertheless, it is clear that a large and unexpected difference between estimates would show that at least one measure is invalid.

Expectations-based validity

14.13 The WTP/WTA values produced by an SP study have **expectations-based validity** if the measures perform in a variety of theoretically sensible ways in relation to other variables, and any non-obvious associations can be explained so as not to cast doubt on the assumed meaning of the measure itself. Expectations-based validity testing is typically achieved through the estimation of a **bid function** relating WTP or WTA responses to a variety of covariates collected in the survey.

14.14 Typically one parameter is estimated for each covariate. This parameter may estimate, for example, by how much a 10% rise in income will affect WTP or WTA. Inspection of these parameters establishes whether variables for which there are prior expectations are both significant in determining valuation responses and affect those responses in the manner expected.

14.15 If crucial variables are found either to be insignificant or, even more importantly, to affect stated values in an unexpected and inexplicable way, this may cast doubt on the theoretical validity of the results. If values are being determined by unexpected factors then this may raise further questions regarding validity or indicate that the processes determining stated values are more complex than realised and/or that these unexpected predictors are proxies for significant omitted variables.

14.16 Alternatively, validity tests will **compare WTP estimates coming from different versions** of the CV questionnaire. For example, analysts may wish to compare WTP values derived from a close-ended elicitation question with those derived from an open-ended elicitation question. Alternatively, they may wish to compare respondents' WTP for a small quantity of the good compared to respondents' WTP for a large quantity of the good.

14.17 Expectation-based testing will frequently be the major form of validity testing conducted on any given study. High quality and appropriate use of data analysis techniques should be demonstrated in the conduct of these tests (see Chapter 12). The relevant expectations are derived from economic theory, prior intuition and observed empirical regularities, and are discussed here. If value estimates are to be used in cost-benefit or similar analyses then expectations derived from economic theory are particularly important. Expectations that should be considered include:

- **Price of the good**. Following a central tenet of economic theory, we would expect that as the price of a good increases then, other things being equal, consumption of that good will fall. This expectation can be tested in all elicitation formats that present subsamples of respondents with different prices and ask whether they are willing to pay this amount or not. A study that does *not* meet this requirement can be said to have failed.

- **Respondent income**. An expectations-based test of validity is that WTP should rise with income, although exceptions may occur. Estimates of the income elasticity of WTP, i.e. how WTP for a fixed quantity of a good changes as income changes, can be derived and compared to those found elsewhere in the literature. This should not be confused with the income elasticity of demand, i.e. how the quantity demanded changes as income changes. Such income elasticities could be below unity, i.e. elasticities may be low, but greater than zero.

- **Quantity of the good**. We would expect, from economic theory, that respondents' WTP for the provision of a good will be related to the quantity of the good provided. Note that we would not expect WTP per unit of the good to be constant regardless of the quantity; in fact we would expect WTP per unit to decline as quantity increases. SP surveys often include tests in order to ascertain whether this expectation holds. Such tests are important in establishing construct validity but should not be considered compulsory. The results of such tests may reveal:

  - **Scope insensitivity,** which occurs when WTP does not vary with the quantity of the good offered. WTP would normally be expected to rise as the quantity of the good on offer increases, but an individual's WTP for incremental amounts of the good would be expected to fall.

- **Embedding** (also known as nesting), which occurs when WTP does not vary between two alternative 'offers', one of which contains a quantity of a good that is also included in the second offer which has another good as well. The first quantity is said to be embedded in the second bundle of goods.

Though such findings may cast doubt on the validity of the CV study, it is possible that scope insensitivity/embedding may simply result from:

- **Satiation**. WTP could stay the same for increasing quantities of a good, i.e. incremental WTP could be zero, if the respondent is satiated in that good. In this case, the insensitivity need not be irrational.

- **Warm glow**. Respondents may get **moral satisfaction** from the act of paying for the good, and this satisfaction may be invariant with the quantity of the good. But note that as shown in table 9.7, bids attributed to the warm glow effect should not be considered as valid.

If such explanations can be justified in the context of a particular SP study then scope insensitivity/embedding may not be an overly burdensome criticism.

- **Sequencing**. Policy makers may be interested in establishing the independent values placed on the component goods provided as part of one policy package. One way of arriving at these component values is to present the respondent with a series or **sequence of goods** in which the policy package is built up incrementally, with valuations being sought at each stage.

14.18 It is possible that WTP varies according to where in a sequence a particular good appears. The later a good appears in a sequence the lower WTP tends to be, and this may appear counter-intuitive. However, evidence of sequencing may not in fact be a bias. If the good appears lower down a list of sequenced goods, disposable income will be less due to the fact that the earlier-sequenced goods have already been 'purchased'. Substitution possibilities may also be greater. Hence economic theory would predict lower values for goods appearing later in a sequence. Since economic theory predicts that the values placed on the different components will be determined by their positioning in the sequence, researchers should avoid sequential design unless the sequence is dictated by the actual policy or programme being valued.

Characteristics of the good

14.19 One of the most basic tests of theoretical validity is to examine whether stated values vary in accordance with the characteristics of the good under consideration. Note that an SP survey values not only the simple physical definition of a good but the wider package which describes the mode of its provision including institutional arrangements, payment vehicle, paying and non-paying populations, who will use the good and who will be excluded, and so on.

14.20 There are no prior expectations about the invariance of results with respect to the wider delivery package relating to a good. We now consider elements of this package that have been known to, or that may be expected to, impact upon reported values, and we discuss how these variations might be exploited for the purposes of validity testing:

- **Scenario.** The credibility and realism of the valuation scenario presented to respondents is crucial to the validity of any SP study. Tests should be undertaken to assess such credibility and examine impacts upon stated values. Because scenario credibility is a cornerstone of the incentive compatibility of any contingent market, any significant problems are liable to induce strategic behaviour and any major problems will undermine the valuation exercise.

  However, more minor variations may provide useful expectations. For example, if some respondents feel that the good described is in some way larger or of higher quality than that which will be delivered, then this should increase stated values for those individuals. Variations, such as the degree to which the relevant providing authorities are trusted or the efficiency or equity of payment collection, may provide useful predictors of variations in stated values.

- **Information effects.** There is a strong body of evidence showing that the provision of information can have an effect upon the values stated in SP surveys. This is as expected and empirical tests indicate that the direction and even magnitude of these effects are typically in accordance with prior expectations. Validity testing (via bid functions and split sample tests) can be affected through variation of the degree of information and inspection of impacts upon stated values.

Characteristics of the respondent

14.21 In any SP study there are usually several expectations regarding the relationship between stated values and factors such as use of the good, the socio-economic characteristics of the respondents and the positive and normative attitudes which they hold. Indeed the assertion that such relationships hold forms the basis of the SP technique. These relationships can be summarised as follows:

- **Use of the good.** The relationship between the use of a good and stated values is expected to be positive and vary directly with the particular degree of use.

- **Attitudes towards good.** Associations between stated values and reported attitudes, membership of interest groups and other indicators of concern regarding a good can be reasonably hypothesised. The testing of such hypotheses is a basic element of theoretical validity analyses.

- **Attitudes towards whole scenario.** Attitudes regarding the equity, fairness, justice and trust dimensions of the valuation scenario will impact upon stated values. Again prior expectations can be formulated and tested, but to the extent that these concerns constitute bias in value estimates these attitudinal tests will simply confirm the presence of such bias.

- **Distance.** For spatially confined goods there are good reasons to expect certain categories of value to decline with increasing distance. Researchers should include an estimate of distance decay within the bid function, i.e. estimate how WTP varies as the respondent's distance from the object in question increases. Distance decay functions are a fundamental element of aggregation procedures (see Chapter 15).

Validity: bias and study design

14.22 In presenting an SP questionnaire to a respondent, researchers are faced by a number of design issues (e.g. to ask for a WTP or a WTA value, whether respondents should be interviewed in person or over the phone etc.). Empirical evidence suggests that values might be influenced by the choice of questionnaire design. A number of these design issues are now considered, and also whether differences in values are evidence of biased valuations and, if so, how such biases might be avoided.

14.23 **Willingness to pay and willingness to accept.** One of the most common results found in the SP literature is that the amount respondents are willing to pay for gaining an increment of a good is considerably less than the amount they are willing to accept in compensation for giving up that same unit of the good. Although this large disparity is often described as **loss aversion**, i.e. losses have greater subjective significance than equal gains, opinion is divided concerning the extent to which such results can be reconciled with standard economic theory.

14.24 One school of thought is that WTA/WTP disparities arise when there are only **limited substitution possibilities** for the good being valued. Such disparities may be particularly noticeable when respondents have a low endowment ('stock') of the good in question. Such an explanation accords with the classical view of preferences. Others contend that preferences are **reference dependent**. That is, preferences may be dependent on some reference point as viewed by the respondent. WTP for an improvement from this reference point — often the status quo — may then be different, and significantly less than WTA to tolerate an equal unit deterioration.

14.25 Which is correct may matter for valuation purposes. For example, reference dependent loss aversion may be specific to *money*, in which case CV would be questionable since it explicitly seeks trade-offs between goods and money. In this case, choice-modelling techniques may fare better since they involve other trades between characteristics of alternative goods.

14.26 Focus groups and pilot studies should test for disparity between WTP and WTA. The resulting final questionnaire might then include both WTP and WTA questions, with some judgement being made about which is more appropriate to the context. A 'first cut' rule would be to use WTA where property rights rest with the potential losers. Others have suggested using WTP due to its 'more conservative' nature and because of a belief that SP has a general hypothetical bias upwards. The evidence for the latter belief is not, however, strong.

14.27 In the context of validity WTP/WTA disparities are likely to be inflated by poor design, low scenario credibility, rejection of the assumed WTA property right and unfamiliarity with the good in question. In effect WTP/WTA disparity should be viewed as both a theoretical problem and a content validity issue.

14.28 **Payment vehicle.** The payment vehicle describes the route through which payment or compensation will hypothetically be made. Researchers typically opt to use the payment

vehicle that would be employed if the programme or policy were implemented in the real world.

14.29 However, the SP literature reports that different values may be returned when respondents are faced with an alternative payment vehicle. Two alternative explanations have been mooted for this observation. First, the way in which a respondent is asked to pay for the good is a substantive part of the package on offer. Accordingly, a respondent's preference for the method of payment may well be reflected in their valuation of that package. Second, an appropriate payment vehicle should be credible, relevant, acceptable, and coercive. If a particular vehicle fails to achieve these criteria then this may induce bias in respondents' valuations. For example, a payment vehicle modelled around a voluntary contribution may well encourage free-riding behaviour and a preponderance of low or zero valuations. Analysts should carefully pre-test alternative payment vehicles in order to ensure they do not conflict with the criterion set out above.

14.30 **Elicitation format.** The different methods available for eliciting WTP and WTA responses in CV studies are described in Chapter 9. It might be argued that WTP should be the same regardless of the elicitation format used (open ended, discrete choice, payment card etc.). In practice, WTP is found to vary with elicitation format. In fact, sensitivity to elicitation format would be expected if the formats differ in the extent to which they embody **incentive compatibility**, i.e. incentives to tell the truth. Binary discrete choice questions are generally expected to produce more truthful responses, explaining the general recommendation to use dichotomous choice and payment card procedures rather than open-ended approaches. However, in any particular study, it is necessary to consider the balance between vulnerability to undesirable effects, the ease with which questions can be administered and understood, and the sample size requirements and other variable costs associated with different elicitation procedures. Those bidding for contracts to undertake SP studies should be expected to address these issues and make explicit their grounds for proposing to use some procedure(s) rather than others.

14.31 **Budget reminders.** It has been recommended that respondents be made to consider budget constraints when answering SP questions and some evidence exists to show that such reminders may influence stated values. Unfortunately, there is little consensus as to whether the budget reminders act so as to reduce or increase respondents' valuations. Thus considerable uncertainty exists regarding the optimum design of such a reminder and this remains a somewhat open question.

14.32 **Survey method.** Theoretical expectations are that there should not be a substantial effect of survey mode (e.g. mailshot, face-to-face interviews etc) on the WTP/WTA results. Survey effects will arise where the choice of mode implies a choice of underlying population but if the population is identical then differences can be viewed as biases.

14.33 **Interviewer effects.** These have been detected in a considerable number of studies. Effects may arise as a result of the interviewer's gender, race, education, appearance, dress, or perceived orientation to the issue. The magnitude of effects is variable but is generally considered to be relatively minor. Interviewer impacts can be reduced through the use of high quality, i.e. consistent and neutral, interviewers and screening out interviewer types thought likely to induce effects. A test for such effects is readily implemented and should be considered as a standard feature of all studies.

Validity and reliability for CM approaches

14.34 In general, tests of validity and reliability will be the same for choice modelling experiments as they are for CV experiments. CM also allows for further tests of validity. These are covered in the Manual,[12] and include:

- Testing internal validity by extending the choice questions beyond those that are of immediate interest, processing the model for the choices of interest, and then seeing if the model 'predicts' the remaining choices;
- Assessing whether respondents are making 'rational' choices by seeing if options which are clearly inferior (e.g. have less of everything) are chosen over superior options; and
- If respondents have been presented with a series of choice experiments, assessing the extent to which the respondents are consistent by comparing their choices over the different choice experiments.

12 I. Bateman *et al.* (2002), *op. cit.*, chapter 8

CHAPTER 15
Aggregation and reporting

Summary

Aggregation is the technique of moving from the sample WTP to estimating the population WTP value. The data may frequently not be ideal, and solutions are provided here for two particular kinds of data problem: unrepresentative samples, and inability to define the relevant population. Clear reporting of project details is important, as ensuring that the methodology is explicit will aid credibility and assist future users of the data. A specification for reporting is provided.

Aggregation

15.01 Chapters 12 and 13 described the steps that need to be taken to produce welfare estimates from data collected in a CV or CM study. These welfare estimates describe the mean or median WTP of the sample of respondents included in the study. In general, however, policy makers are not interested in the mean or median WTP of the **sample** but in the mean or median WTP of the relevant **population**. The final task is therefore to aggregate the values which are of interest from the sample to the population.

15.02 This task is typically undertaken at the very end of an SP exercise. However, questionnaires must be designed and sampling procedures executed in specific ways in order that successful and valid aggregation measures can be calculated. Indeed, many of the problems of aggregation can only be avoided by carefully selecting the sample. Hence much of the focus is again on the **design of a sampling strategy** (see Chapter 8).

Aggregation in an ideal study: the simple random sample

15.03 Ideally, an SP study would proceed through the following five steps:

1. the **population of interest** is chosen (e.g. the users of a given recreational site);
2. the **unit of observation** is chosen (e.g. the visiting household);
3. a **simple random sample** of those units is drawn, with each unit in the population having the same positive probability of inclusion in the sample;
4. all units chosen for the sample agree to be interviewed and provide complete responses to all questions. That is, there is **no non-response** to the survey, and

5. the **statistic(s) of interest is(are) estimated** for the sample (e.g. mean WTP is chosen and an unbiased estimate is obtained).

15.04 In this case the aggregation process is simple. If we denote the statistic of interest (often the sample mean or median WTP) as $\hat{WTP}$ and the total number of units in the population as N, then:

$$\text{Aggregate WTP} = N \cdot \hat{WTP}$$

That is, aggregate $\hat{WTP}$ can be calculated simply by multiplying $\hat{WTP}$ by the number of people in the population.

15.05 However, it is frequently difficult or even impossible to ensure that steps 1 to 5 are followed faithfully, so that more complex issues arise in the aggregation process. Hence, much of the art of aggregation concerns strategies for dealing with failures to meet all of the above conditions. In particular, digressions from the ideal study usually relate to an **unrepresentative sample** (failure in steps 3 and 4) and an **inability to define the relevant aggregating population** (failure in step 1).

Unrepresentative samples

15.06 Assuming that the population of interest is known but that the analyst is faced with data representing a non-random sample of that population, two situations then suggest themselves.

15.07 **Probabilistic samples.** A relatively simple digression occurs where the sample is probabilistic but not perfectly random. That is, a random sample of units has been drawn from the population with each unit in the population having a known and positive, though not equal, probability of inclusion in the sample.

15.08 The analyst can use the probability of inclusion in the sample to define an analytical weight for each observation i, that is denoted w_i (see Chapter 12). If these weights are used in the estimation of $\hat{WTP}$, then the above formula applies. Alternatively, the analyst can define $\hat{WTP}_i$ for each observation in the sample of n observations. If the data are collected using an open-ended question format then $\hat{WTP}_i = WTP_i$, the stated WTP of each respondent. If the data are collected using a different elicitation format then $\hat{WTP}_i$ can be estimated from the bid function. The aggregation formula is then modified to:

$$\text{Aggregate WTP} = N \cdot \sum_{i=1}^{n} w_i \hat{WTP}_i$$

15.09 **Non-probabilistic samples.** If sampling has been non-probabilistic or the characteristics of the sample have been biased as a result of non-response, then the specific probability of a household being in the sample is difficult to determine. Fortunately, the analyst can retrospectively define analytical weights by comparing the characteristics of the households represented in the sample with those of the population in general.

15.10 As described in Chapter 12, weights can be defined by dividing the proportion of the population falling into a particular characteristic grouping by the proportion of the sample falling into that same group. Again the weights can be used in the estimation of $\hat{WTP}$ in which case the first aggregation equation is used. Alternatively, the weights can be used

directly in the aggregation process in which case the second aggregation equation is the relevant one.

Defining the aggregating population

15.11 In some circumstances the analyst can be fairly certain that the sample is representative but may be unable to define the relevant population. Such a scenario might occur with a random intercept sample of visitors to a recreational area: the sample is random but the researcher does not know the total population of users. In this case there are three options:

- conduct additional research to estimate the total population;
- guess at the total population, or
- arbitrarily define the population as belonging to some administrative area.

15.12 Clearly the latter two options are far from ideal approximations, and where possible analysts should seek to carry out further research in order to arrive at an estimate of the total population. Once an estimate of the total population is available then aggregation can proceed using the first aggregation equation.

15.13 The survey data frequently become a useful and sometimes optimal source of information regarding the true aggregation population. One simple case is where the good being valued has a spatial dimension. The analyst can partition the population into zones differing in distance from the site of provision. The WTP of each household in that zone can be taken as the mean WTP of individuals sampled in that zone. For zones over a certain distance, it is assumed that WTP equals zero. The relevant distance can either be identified from the data or assumed by the analyst. If WTP is found to be invariant with distance, then the correct procedure is to aggregate across the whole national population.

15.14 A variant on this approach is to estimate a bid function containing distance from the site of provision of the good as a covariate (see Chapter 12). It should then be possible to see how WTP varies with distance from the site, so as to establish a **distance-decay** function. The distance-decay function can be used to identify the distance at which WTP is zero. The WTP of all households within this limit can be calculated from the function. Summing these values will give aggregate WTP. Further generalisations of this theme can be adopted. The bid function can be estimated in two parts:

- **the participation function,** which estimates the probability of a household having a positive WTP. Note that the participation function may also account for the likelihood of the household participating in the SP survey in the first place, and
- **the bid function,** which estimates the WTP of those with positive WTP.

15.15 Both these functions can be made dependent on distance and other socio-economic variables. Using these two functions, reasonably sophisticated aggregation can be undertaken. For example, for a recreational site, an analyst might undertake the following steps:

- define a geographical region that is assumed to include all those who value the recreational area;

- divide this area into a series of zones (the smaller the zones the greater the resolution of the analysis);
- for each zone extract information on population and income from national statistics and calculate the distance from the zone to the recreational site;
- on the basis of a geographically distributed sample survey, estimate the participation function as a function of distance and income;
- feed the zonal income and distance data into the participation function to calculate the predicted number of households in each zone with a positive WTP;
- using the whole sample, estimate a bid function as a function of income and distance from the site;
- for each zone, substitute the level of income and distance to the site in the bid function to obtain mean WTP by household;
- multiply WTP per household for a given zone by the estimated number of households in that zone with WTP greater than zero to obtain aggregate WTP for that zone, and
- aggregate each zonal aggregate WTP across zones to give total WTP.

Reporting

15.16 Reporting should be transparent and comprehensive, so that readers can obtain as much information as possible about what was done and why. This will prevent misunderstandings about the approaches used, and should enhance the credibility of the results. Full information will also make it easier either to replicate the methodology in other contexts, or to use the results for benefits transfer. The report should contain at least the following sections.

- **Executive summary** containing the main findings and describing the broad approach, both in non-technical language.
- **Description of the objectives of the study,** including a detailed description of the non-market effect being valued; descriptions of attributes of the non-market effect that might vary in a final programme or policy and other relevant information concerning the attitudes or opinions of the population that might usefully be collected as part of a survey.
- **Review of any previous relevant valuation studies,** reporting on the choice of valuation techniques (e.g revealed preference, contingent valuation, choice modelling); an outline of the design of previous SP surveys, indicating relevant information on the choice of scenario, payment vehicle and institutional context used in the survey instrument; difficulties and problems encountered with these designs, highlighting lessons to be taken on in the current research and values derived from the studies.

- **Description of the process of survey design**. Drawing on the results of focus groups, pre-tests, previous research and peer reviews of the study, the report should cover the following:

Overall approach: How the decision to employ CV or CM was reached; why, if at all, split samples have been used; and how the attributes of the non-market effect have been chosen and measured in a CM study.

Issues of scenario design: Whether or why the unit of observation (e.g. household or individual) is suitable for the good being valued; assurance that the good offered is specified to and understood by respondents; adequacy of the information provided to respondents and the description of the change and payment scenario; the plausibility of the trade-off between money and the good; and assurance that descriptions of substitutes and the consequences of non-payment are adequately described and that respondents were adequately reminded of their budget constraints.

Elicitation issues: whether the chosen measure of wellbeing is appropriate (WTP or WTA) and whether the chosen elicitation format (again, WTP or WTA) is appropriate.

Institutional context: whether the method of provision and allied institutional arrangements are plausible; that respondents are likely to have an expectation of having to pay for the good if it is provided and that respondents are likely to feel that they are providing an input to the decision-making process.

- **Description of the survey administration procedure,** reporting on the date and location of the interviews; the definition of the population of interest from which the sample has been drawn; the choice of sampling frame (e.g. simple random sample, stratified probabilistic sample etc.); the choice of survey mode (e.g. face to face interviews, mail surveys etc.), the levels of unit non-response and the representativeness of the final sample.

- **Summary of data,** providing an overview of the general picture of respondents, numbers, sample percentage, etc to provide a picture of the population. This can be important, depending on which population is targeted.

- **Details of the data analysis (CV):** reporting on the identification of non-valid responses and stating how many responses have been categorised as non-valid and for what reason; a non-parametric model used to derive lower bound estimates of mean and median WTP/WTA; a constant only bid function model from which a 'best' mean and median WTP/WTA estimates are derived; confidence intervals for the estimates of mean and median WTP; a fully parametric model used to estimate a bid function that includes a variety of variables such as income, socio-economic details and possibly distance to the site of provision; a measure of the overall fit of the model; tabulation of the parameter estimates, their standard errors and statistical significance; a discussion of the parameter estimates examining whether they conform with prior expectations and a benefit transfer model estimated including only basic socio-economic variables.

- **Details of the data analysis (CM):** reporting on the identification of non-valid responses reporting how many and for what reason responses have been categorised as non-valid; the description of the choice of random utility model; tests for independence of irrelevant alternatives in conditional logit models; a measure of the overall fit of the model; tabulation of the parameter estimates, their standard errors and statistical significance; a discussion of the parameter estimates examining whether they conform

with prior expectations; welfare estimates for defined programmes and implicit prices and confidence intervals for the welfare estimates.

- **Tests of validity** reporting on the indications of low content validity and the indications of construct validity.

- **The process of aggregation** reporting on how sample values of WTP have been aggregated to the population and final values for aggregate WTP.

- **A copy of the questionnaire** (and any split sample versions) should be included as an Annex to the report.

Conclusion

15.17 In conclusion, the use of stated preference techniques is a complex but increasingly used means of establishing money values for impacts which do not themselves have observable money values. This summary guide shows that substantial professional help will be required in successfully conducting and analysing any stated preference survey in a way that avoids most of the pitfalls and produces reliable and robust results. The techniques discussed here are constantly being developed and refined, and practices that were at one time seen as very advanced have tended to become essential for the sake of credible results which can support policy decisions. The key message of this text is that SP approaches have an important role to play in providing valuation data for decision making, but they have to be applied with methodological rigour and with adequate resources to produce a good job.