

## REVIEW OF STATED PREFERENCE AND WILLINGNESS TO PAY METHODS

### Introductory note by the Competition Commission

1. The Competition Commission (CC) commissioned the market research organization, Accent, in association with RAND Europe, to conduct a review of 'Stated Preference and Willingness to Pay Methods' in conducting surveys. The review report was completed in April 2010, and is published below.

### Background

2. The CC frequently uses surveys when conducting competition investigations and the evidence from these surveys is an important component of its findings. Survey evidence is also proving to be useful in remedies work.
3. This report (attached below), by Accent and RAND Europe, covers the methodological options for carrying out stated preference studies; sets out how the studies should be designed and sets out the market research methods to conduct the studies.

### How the study was undertaken

4. The study consisted of four elements:
  - a workshop with the study team and CC representatives;
  - literature review;
  - telephone interviews with representatives of three overseas competition authorities; and
  - review of five recent CC investigations.<sup>1</sup>

### Outputs

5. The study covered four main areas:
  - choosing willingness to pay (WTP) methods;
  - process for designing stated preference research;
  - appraisal of stated preference valuation techniques in practical applications; and
  - review of current practices.

### *Choosing willingness to pay*

6. At the highest level the literature classifies the different methods for estimating WTP into revealed and stated preference. The first of these refers to the observation of

---

<sup>1</sup>Four of the five are transport studies.

preferences revealed by actual market behaviour. In cases where it may not be possible to observe such preferences it is necessary to make judgements about potential impacts, in the absence of real-world evidence on how individual consumers may respond. Stated preference (SP) methods allow such examination. The report focuses on these SP methods.

7. SP methods include both non-use value and option value and this widens their use as they can be used to value potential future or hypothetical goods and interventions. There are two main categories of SP methods: contingent valuation methods (CVM) and choice modelling techniques, which includes discrete choice experiments (DCE)

### ***Process for designing stated preference research***

8. The report reviews the potential CVM approach versions that can be used and concludes that, overall, payment cards and dichotomous choice formats are the most recommended CVM elicitation methods. CVM uses direct question to elicit customers WTP, for example, 'What are you willing to pay for this service?'.
  9. The key steps in designing a CVM questionnaire are:
    - formulating the valuation problem;
    - drafting additional questions (debriefing, attitudes and demographics); and
    - pre-testing of the questionnaire.
  10. DCE, a choice modelling technique, presents customers with alternative scenarios to choose between. The report shows that among the design steps used to generate a choice modelling experiment the following are key steps:
    - defining the problem;
    - carrying out a qualitative study (to identify alternatives, attributes and levels);
    - designing the experiment;
    - generating choice sets; and
    - constructing survey instruments.
  11. Enhancing SP information by combining it with revealed preference data is discussed by setting out both the advantages of the combination, and the challenges that may arise.

### ***Appraisal of stated preference valuation techniques in practical applications***

12. When deciding between CVM and DCE the analyst has to strike a balance between the rigour of the method and the time and budget available. The key assessment issues are:
  - cost of survey: CVM lower than DCE;
  - timescale: CVM shorter than DCE;
  - valuations: DCE can value individual attributes, CVM the total package only;

- complexity of design: CVM less complex than DCE;
- software and analysts: DCE requires specialist software and analysts, CVM does not;
- complexity of task for respondents: more complex in DCE than CVM;
- compliance bias: high in CVM, low in DCE;
- stability of preferences: better in DCE than CVM; and
- simultaneous estimation of marginal effects and attribute values: in DCE but not CVM.

### ***Review of current practice***

13. The report reviews five recent CC investigations which focused on cases involving competition at the local or national level where the hypothetical questions asked were in the form of contingent valuation questions or discrete choice experiments.

### **Recommendations**

14. The report makes recommendations with respect to the design of CVM and DCE survey questions and on more general market research issues.

### ***Design of CVM and DCE***

15. The report recommends the use of DCE when carrying out WTP research. However, the researchers stated that the time pressures within which the CC works would often lead to CVM being a more likely instrument of choice. When undertaking future surveys actions worth considering are to:
  - include an additional stage to inform respondents of current alternative prices and then give them the opportunity to indicate any change in behaviour with this knowledge;
  - ensure that questions are clear and as intended (ie with no ambiguity);
  - adopt a simple experimental design when a number of alternatives are being considered;
  - introduce any choice valuation task by setting the context for the issue whilst reminding the respondents that this falls within their wider day-to-day considerations (ie 'imagine you face this situation where...'); and
  - include diagnostic questions to assess the extent to which the respondents understood the task.

### ***Market research***

16. The report also looks at the wider area of market research within which stated preference work will be carried out and gives some recommendations on how to carry out such research:

- conduct pilots wherever possible;
- use qualitative research to help identify marginal customers;
- ensure the sample is random, accepting that quota sampling maybe necessary in certain cases;
- ensure a robust sample size—400 for overall sample and 75 in a cell for stated preference;
- use a short questionnaire—keep to 10 minutes where possible and no longer than 30 minutes even when using DCE;
- mention the CC at the beginning of the survey;
- note refusals, if possible, to ensure that the sample is representative of the population;
- minimize non-response error by maximizing response rates; and
- use ‘cheap talk’<sup>2</sup> approach to ensure understanding.

### **Future use of this work**

17. The CC often commissions consumer surveys to elicit customers’ WTP and such evidence has been used to inform the decision making process in many recent investigations. As for all techniques there are always new methods to consider and learn and the CC is keen to ensure that this continued development takes place in the area of consumer survey design. With this in mind, any suggestions on ways to improve the quality of information collected are extremely welcome.
18. This report will therefore help the setting up of stated preference exercises in the future. This report should not be considered best practice, or to be specific guidelines from the CC on how to do stated preference exercises, but a useful set of recommendations that may be applied in such surveys.
19. Recent surveys undertaken by the CC have already incorporated learnings from this study in the set up of discrete choice experiments and the CC anticipates using these learnings in future inquiry work as the need and circumstance arise, taking account of the time to carry out such techniques.
20. The full text of the report follows below.

Marcia Stephens  
*Senior Statistician*  
 October 2010

---

<sup>2</sup>‘Cheap talk’ alerts the respondents to the issue of hypothetical bias just before asking the hypothetical questions to try to minimize this form of bias.



## Review of Stated Preference and Willingness to Pay Methods

**April 2010**

Prepared by:

Accent  
Chiswick Gate  
598-608 Chiswick High Road  
London  
W4 5RT

Prepared for:

Competition Commission  
Victoria House  
Southampton Row  
London  
WC1B 4AD

*In association with:*

RAND Europe

---

Contact: Chris Heywood  
Tel: 020 8742 2211  
E-mail: [chris.heywood@accent-mr.com](mailto:chris.heywood@accent-mr.com)

Contact: Marcia Stephens

File name: 2003rep06.doc

# CONTENTS

Executive Summary .....	i
1. INTRODUCTION .....	1
1.1 Background.....	1
1.2 Report Structure.....	1
1.3 Terminology and Its Use across Sectors .....	1
1.4 List of Abbreviations and Acronyms .....	2
2. CHOOSING WILLINGNESS-TO-PAY METHODS .....	3
2.1 Introduction .....	3
2.2 Revealed Preference Methods .....	5
2.3 Stated Preference Methods .....	7
3. DESIGNING STATED PREFERENCE EXPERIMENTS .....	14
3.1 Introduction .....	14
3.2 Designing Contingent Valuation Questions .....	14
3.3 Designing Choice Modelling Experiments .....	22
3.4 General Design Considerations .....	32
3.5 Enhancing SP Information: Combining Stated with Revealed Preference Data.....	37
4. APPRAISAL OF SP VALUATION TECHNIQUES IN PRACTICAL APPLICATIONS.....	40
5. REVIEW OF THE CURRENT PRACTICE.....	44
5.1 Introduction .....	44
5.2 Review of Pricing Questions Used by the Competition Commission.....	44
5.3 Moving from CVM to DCE: An illustration using the Eastman Kodak Company (2001) study .....	54
5.4 Recommendations from Review of Current Competition Commission surveys .....	61
5.5 Market Research Methodology Recommendations.....	62
6. REFERENCES .....	65
Appendix A: International Depths Topic Guide	
Appendix B: Other Survey Issues	
Appendix C: Current Practice in Competition Commission Surveys	
Appendix D: Current Practice in Other Competition Authorities	

# EXECUTIVE SUMMARY

## Why study Commissioned

The Competition Commission (CC) frequently uses surveys when conducting competition investigations and the evidence from these surveys are an important component of its findings. Survey evidence is also proving to be useful in remedies work.

Such surveys are undertaken in a wide range of sectors which has included, for example, personal banking, gaming, betting, retail, travel, packaging, advertising and storecards.

They are commissioned to fill in gaps in knowledge such as customer response to price changes, customer switching behaviour if there are changes in the market and the value that customers put on alternatives and their attributes. Such research can inform market definition and the analysis of competitive effects.

This report covers the methodological options for carrying out stated preference studies; sets out how the studies should be designed and sets out the market research methods to conduct the studies.

## How Study was Undertaken

The study involved the following stages:

- Workshop with the study team and CC representatives. This was designed to understand the CC's requirements from the hypothetical research questions and their particular concerns with current approaches and stated preference methods
- Literature review on SP and WTP methods
- Telephone interviews with representatives of the US Department of Justice, the EU Directorate General for Competition and the US Federal Trade Commission
- Review of five recent CC studies in order to explore how the current practice of CC compares with the current practice reported in the academic literature and other competition commissions worldwide.

## Study Outputs

The study covered the following areas:

- Choosing willingness-to-pay methods
- Process for designing stated preference research
- Appraisal of stated preference valuation techniques in practical applications
- Review of the current practice

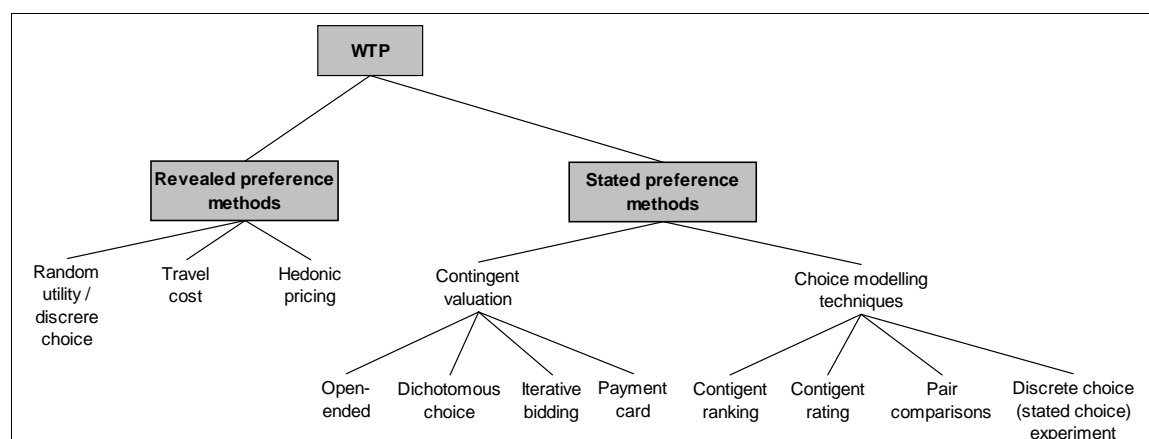
Each of these are discussed in turn.

## Choosing Willingness-To-Pay Methods

A key area of competition analysis is an assessment of how the public may be likely to respond to changes in the services which are available to them, particularly the value that they place upon different aspects of the services available, and how their behaviour may change if the alternatives available were to change.

In many situations the analyst would wish to look to the revealed trends in market behaviour though this is not always feasible. As a result, a significant literature has been developed around survey methods for estimating individuals' willingness-to-pay (WTP) in the absence of revealed market variation. These methods are now widely used for both developing optimal pricing strategies and also in the forecasting of responses to price changes and for modelling demand functions.

At the highest level, the literature classifies the different methods for estimating WTP into revealed and stated preference methods as shown in the figure below.



**Figure E1: Willingness-to-pay methods (sources: Bateman et al. (2002), Kjaer (2005))**

Revealed preference methods (RP) refer to the observation of preferences revealed by actual market behaviour and represents real-world evidence on the choices that individuals exercise.

In some cases, however, the behaviour that is of interest to the analyst may not be observable or currently available. In such cases it is necessary to make judgements about potential impacts in the absence of real-world evidence on how individual consumers may respond. Stated preference (SP) methods allow examination of such hypothetical situations.

To date, the Competition Commission's willingness to pay studies have been largely based on stated preference data, mainly because revealed preference (actual choices) data are unavailable or markets do not currently exist. Therefore, the report focuses on stated preference approaches.

## **Stated Preference Methods**

The family of stated preference (SP) methods can measure the total economic value; that is, SP methods incorporate both non-use value and option value. This characteristic has far-reaching potential as it implies that SP can be used to value potential future or hypothetical (but realistic) goods and interventions. When considering stated preference methods, the main categories are the Contingent Valuation Methods (CVM) and choice modelling techniques including Discrete Choice Experiments (DCE).

Both approaches have their merits but choice modelling approaches appear to be in the ascendancy as they provide a more direct route to the valuation of the *characteristics* or *attributes* of a good, and of marginal changes in these characteristics, rather than the value of the good as a whole. This may be of particular importance in management decisions, project appraisals and policy appraisals where the decision is sometimes based on changes in the levels that these attributes take. For example, changes in the availability of services may require users to travel farther and use another service with different level of quality and price characteristics

## **Choice Modelling Techniques**

Within the choice modelling family there are a number of different techniques. In principle, all choice modelling techniques assume that goods or services can be described in terms of their attributes or characteristics and the levels that these take. The focus is on the value placed on these attributes.

Choice modelling techniques can be classified into four main categories, which reflect differences in theoretical assumptions, methods of analysis and experimental design procedures.

In terms of practical application, the academic literature over recent years, however, has largely converged around discrete choice (or stated choice) experiments (DCE) as it provides a task that most closely mirrors that which people face in real life (ie a single choice between a set of alternatives). Moreover, DCE allows for indirect estimation of WTP, which is the recommended approach by the HM Treasury for the valuation of changes in quantity or public services. On the other hand, Best-Worst Scaling is a useful method for estimating the values (not preferences) placed on certain attributes on a common interval scale, but it cannot be used to estimate future demand and elasticities.

## **Process for designing stated preference research**

The report reviews the potential CVM approach versions that can be used and concludes that, overall, payment cards and dichotomous choice formats are the most recommended CVM elicitation methods. Payment cards offer more information and less cognitive burden to respondents and are cheaper to implement than dichotomous choice. In addition they are superior to both direct open-ended questions and bidding games. Dichotomous choice formats may be incentive compatible and facilitate respondents' valuation task.

## **Designing CVM questionnaires**

The objective of a contingent valuation (CVM) questionnaire is to elicit preferences in monetary terms, more specifically the maximum WTP or minimum Willingness to Accept (WTA) for changes in the quantity or quality of a good or service from a random sample of respondents. These changes may refer to a hypothetical or an actual good or service.

Firstly, CV questionnaires ask respondents to consider how a change in a good or service might affect them. The policy change (real or hypothetical) should be perceived as being realistic and feasible and is described in detail before respondents are asked to evaluate it. Secondly, the hypothetical good or policy may be complex and unfamiliar to respondents which may result in inconsistent and unreliable observations. Finally, respondents are asked to state their WTP or WTA for the change in the good or policy in question.

The review shows the design steps used to design a CVM questionnaire. The key steps are summarised below.

- formulating the valuation problem
- additional questions and
- pre-testing of the questionnaire.

## **Designing Discrete Choice Experiments (DCE)**

The various categories in choice modelling approaches exhibit a number of fundamental differences in both the theoretical assumptions and the structure of the choice tasks. However, a number of common assumptions and design stages do exist. In particular, all choice modelling experiments are based on data generated through a systematic and planned experimental design process in which the attributes and their levels are pre-defined and varied to create preference or choice alternatives.

The review shows the design steps used to generate a choice modelling experiment. The key steps are summarised below.

- Problem definition
- Qualitative study (to identify alternatives, attributes and levels)
- Experimental design
- Generate choice sets
- Construct survey instrument.

## **Uncertainty questions and cheap talk**

The DCE approach is less prone to hypothetical biases in WTP estimates than CVM as it replicates actual consumer purchasing condition more closely than CVM. However, testing hypothetical biases and the recent influence of cheap talk scripts in ‘grounding’ WTP estimates of DCE has been limited to only a few studies.

## **Enhancing SP Information: Combining Stated with Revealed Preference Data**

This review discusses the fact that revealed and stated preference approaches have both advantages and weaknesses. And, historically, researchers have seen revealed and stated preferences as substitutes when considering the choice of valuation methods. Empirical evidence about the validity of stated preference methods and the fact that the strengths of revealed preferences are the weaknesses of stated preference approaches have led to a new paradigm for combining and jointly estimating revealed and stated preference data.

While the combination of RP and SP data presents a number of advantages over the use of RP or SP data alone, a number of challenges include:

- Model complexity and availability of user-friendly software to implement the combination
- How to reconcile potential differences between stated behaviour and actual behaviour in developing predictive models for new scenarios
- Context-specific applicability, as there might exist an RP technique that provides the solution to the problem
- Empirical evidence in the estimation of combined RP and SP models is limited.
- Longer and intensive questionnaires may result in lower response rates, lower quality of responses, etc.

## **Appraisal of SP Valuation Techniques in Practical Applications**

When deciding between CVM and DCE in practical applications, the analyst has to balance between the rigour of the method and the time and budget available. DCE take much longer to implement as they require background study, focus groups and testing of the potential attributes and levels prior to the main survey. The experimental design stage is the most important task in the design of DCE and poses more challenges than the procedure of designing valuation questions in CVM. At the conceptual stage, therefore, the time and budget available are the two most important factors in deciding whether a CVM or a DCE would be preferable.

The ability in DCE to present the good or service in question within a ‘package’ of potential trade-offs is also seen to be a way of minimising information transfer and other potential biases, and modelling realistic choices. This is very important when the good in question is not familiar to respondents. Therefore, the DCE approach relies more on the accuracy and completeness of the characteristics used to describe the good or service than any particular description of the good or service as a whole. Moreover, cost in DCE is just one of the several attributes presented to describe the good or service in question. Thus, the emphasis on cost is less and CVM related biases may be reduced. Overall, the DCE is the recommended approach in cases which valuation of one or more attributes of a good or service are of particular interest.

The key assessment issues when choosing between CVM and DCE are set out below.

- Cost of survey: CVM lower cost than DCE

- Timescale: CVM shorter than DCE
- Valuations: DCE can value individual attributes, CVM total package only
- Complexity of design: CVM less complex than DCE
- Software and analysts: DCE requires specialist software and analysts, CVM does not
- Complexity of task for respondents: more complex in DCE than CVM
- Compliance bias: high in CVM, low in DCE
- Stability of preferences: better in DCE than CVM
- Estimation of marginal effects and attribute values simultaneously: in DCE but not CVM.

## Review of the Current Practice

We reviewed a number of previous CC studies to look at the key patterns. The hypothetical questions in CC's studies were either in the form of contingent valuation questions or discrete choice experiments. Contingent valuation questions usually were developed using a mix of methods (eg dichotomous choice and iterative bidding) whereas discrete choice experiments were developed using standard SP methodology.

All studies focused on hypothetical cases involving potential change in competition at the local or national level, and therefore of particular interest were changes in the behaviour of customers with regard to changes in the supply of services or products. Prices and quality characteristics of products or services were of particular interest.

## Recommendations

The study makes recommendations with respect to the design of CVM and DCE survey and on more general market research issues. These are shown in turn below.

### Design of CVM and DCE surveys

As CC studies have to be delivered within short timeframes, it is inevitable that the CVM type of questions is often the preferred option. However, on the basis of our review of a subset of previous CC surveys we recommend consideration of the following points when undertaking future surveys:

- The analyst should explicitly consider:
  - the extent to which they wish to explore the respondents' current understanding of the choices that they have available
  - the extent to which any switching may be driven by changes in price (or service)
  - whether there is a risk that the survey instrument may confound the response to a price signal as this can result in a potential over-estimate of the price elasticity
  - the framing of the survey question to ensure it does not provide pre-defined response options which the respondent may or may not have perceived as viable, or considered as available alternatives. This can also result in a potential over-estimate of the price elasticity.

- It can be helpful to include an additional stage within the questioning, where the respondent is informed of the current prices of possible alternatives and given an opportunity to indicate a change in behaviour in response to this improved information prior to varying the prices.
- The analyst should also take care to ensure that the question being asked is the question intended (ie with no ambiguity).
- In cases where the survey seeks to examine situations where attributes may be varying across a number of alternatives it is worth considering adopting a simple experimental design, which allows the pricing scenarios being presented to be defined in a manner that will allow the impact of the prices on each of the presented alternatives to be understood independently from each other.
- In the wording used to introduce the choice or valuation task it is useful to set the context for the issue under consideration whilst reminding the respondent that this falls within their wider day-to-day considerations (eg time or budget constraints). It is better to avoid wording such as “in your view ...” which invites a value based judgement and rather use wording such as “imagine that you face the situation where ...” which places the respondent within the context of a more routine choice situation.
- There is benefit from including diagnostic questions at the end of a survey to explore the extent to which the respondents understood any tasks that they had been asked to undertake or to give an opportunity to voice issues which they wanted to raise but felt that the survey questions did not give them opportunity to express. Such data can be useful in cleaning a data set and identifying those that may not have fully understood the questions.

## Market Research Methodology Recommendations

This section focuses on our recommendations for the market research aspects of CC studies.

### Ways of dealing with bias

It is important to minimise the following forms of bias when selecting the sample:

- **sampling error** – the sample not being representative of the population as a whole. Sampling error is minimised by ensuring the sample is randomly selected but this is not always practical for reasons of time, cost and the nature of the sample. We recommend that the CC uses quota samples for their surveys whilst making every effort to ensure that the sampling approach is as random as possible.
- **non-response error** – those not responding being different from those that do. Non response error can be minimised by maximising response rates. We recommend that this is done through the following:
  - using short well designed questionnaires
  - using well trained interviewing staff
  - giving assurances of confidentiality and data protection
  - mentioning the survey sponsor’s name

- using of incentives (if the questionnaire is long and/or complex).

### **Cheap talk**

We recommend that the ‘cheap talk’ approach is considered to correct for hypothetical bias in DCE and CVM surveys in particular. The idea behind the ‘cheap talk’ approach is that a *cheap talk script* makes respondents aware of the problem of hypothetical bias prior to the actual hypothetical questions.

### **Use of CC's name**

We recommend that CC’s name is mentioned in their surveys as the survey sponsor to increase the response rate. We do not consider that mentioning the CC as a survey sponsor will cause any bias

### **Length of questionnaire**

It is important to ensure that the questionnaire is not too long as longer questionnaires tend to lower responses rates and make the survey less representative as well as increasing costs. Incentives can help alleviate concerns with respect to response rates for longer interviews.

The recommended maximum length of the question is partly dependent on research methodology with, for example, longer questionnaires being possible for in home interviews and shorter questionnaires required for *in situ* interviews.

It should also be noted that DCE questionnaires are typically at least 20 minutes in length.

We recommend that CC questionnaires are restricted to 10 minutes in duration where practical and are no longer than 30 minutes even if using DCE.

### **Marginal customers**

Our recommendation with respect to identifying marginal customers (ie respondents most likely to switch behaviour) is to undertake qualitative research to help identify the nature of such respondents and to help define the appropriate question or questions needed for such an identification.

Our recommendations on scaling marginal customers’ responses to the overall population is to ensure that non marginal customers are also be sampled in the research and to ensure that the proportion of marginal customers to all customers is correctly identified so that weighting of the data can be undertaken.

### **Sample size**

As a larger sample size leads to more robust data there is always a trade off in market research between cost and precision. In this respect, statistical reliability is a relative term. We recommend using the market research common practice of using a minimum

total sample size of 400 for simple surveys as this represents a reasonable balance between robustness of results and cost of fieldwork.

However, it needs to be noted that the sample size for each specific sub group of interest needs to be large enough to provide for robust data taking account of the statistical significance when comparing data between groups.

Finally, for stated preference surveys it is recommended that the size of each cell of interest is at least 75.

### **Pilot**

The pilot is an ideal that should be used whenever possible as part of the research process, particularly for stated preference approaches. It is used to test both the survey methodology and the questionnaire.

The pilot sample size for stated preference research should be at least 50.

### **Out of Scopes/Refusals**

We recommend that for all surveys the number of people screened and found to be out of scope and the number who refuse to take part are recorded, so that an estimate of the size of the overall population can be made and to establish whether refusals are representative of the sampled population.

# 1. INTRODUCTION

## 1.1 Background

The Competition Commission (CC) frequently uses surveys when conducting competition investigations and the evidence from these surveys are an important component of its findings. Survey evidence is also proving to be useful in remedies work.

Such surveys are undertaken in a wide range of sectors which has included, for example, personal banking, gaming, betting, retail, travel, packaging, advertising and storecards.

They are commissioned to fill in gaps in knowledge such as customer response to price changes, customer switching behaviour if there are changes in the market and the value that customers put on alternatives and their attributes. Such research can inform market definition and the analysis of competitive effects.

## 1.2 Report Structure

Chapter 2 discusses the different Willingness-To-Pay methods available.

Chapter 3 describes the process for designing stated preference research.

Chapter 4 contains an appraisal of sp valuation techniques in practical applications.

Chapter 5 reviews the current practice adopted by the Competition Commission and includes recommendations.

Chapter 6 includes references.

The appendices include the Topic Guide used for the International Depths (Appendix A), a discussion of market research issues (Appendix B) and a discussion of the current market research practice in Competition Commission surveys (Appendix C) and other competition authorities.

The terminology used across sectors and a list of abbreviations and acronyms are shown in the next two sections.

## 1.3 Terminology and Its Use across Sectors

Term	Definition
Attribute, factor or treatment	A policy element or a characteristic of a product or service such as price, waiting time, tax discount, etc.
Choice set	A finite set of a number of alternatives available to an individual

Term	Definition
Choice task or choice exercise or exercise or card, scenario, profile, alternative, option or treatment combination	The options that are considered by the individual to make the choice. These are described by a series of attributes eg different hospitals will have different waiting times, reputations, etc.
Stated Preference Methods or Stated Preference	<p>In this report, these terms are used to describe the whole family of methods including contingent valuation method (CVM), conjoint analysis and discrete choice (or stated choice) experiment (DCE) methods.</p> <ul style="list-style-type: none"> <li>- <i>Referendum conjoint</i> analysis is considered part of the discrete choice experiment approach when the experiment includes only one attribute</li> <li>- <i>Conjoint analysis</i> is a term introduced by Green and Rao (1971) to refer to any stated preference elicitation method that combines attributes levels into profiles and seeks some type of preference (ie choice, ranking or rating) to the profiles when these are introduced to a sample of individuals in the context of a survey.</li> <li>- <i>Open-ended</i> contingent valuation = ranking (Bateman et al. 2002)</li> <li>- <i>Discrete choice experiment</i> is also known as the stated choice experiment approach</li> </ul>

#### 1.4 List of Abbreviations and Acronyms

BWS	Best-worst Scaling
CA	Conjoint Analysis
CAPI	Computer Aided Personal Interviews
CATI	Computer Aided Telephone Interviews
CC	Competition Commission
CV	Contingent Valuation
CVM	Contingent Valuation Method
DBDC	Double-bounded Dichotomous Choice
DC	Dichotomous Choice
DCE	Discrete Choice Experiment
IB	Iterative Bidding
NOAA	National Oceanic and Atmospheric Administration
OE	Open-ended
PC	Payment Card
PDA	Personal Digital Assistant
RDD	Random Digit Dialing
RP	Revealed Preference
RUT	Random Utility Theory
SC	Stated Choice
SP	Stated Preference
SSNIP	Small but Significant and Non-transitory Increase in Price
TC	Travel Cost
WTA	Willingness-to-accept
WTP	Willingness-to-pay

## 2. CHOOSING WILLINGNESS-TO-PAY METHODS

### 2.1 Introduction

A key area of competition analysis is an assessment of how the public may be likely to respond to changes in the services which are available to them, particularly the value that they place upon different aspects of the services available, and how their behaviour may change if the alternatives available were to change.

In many situations the analyst would wish to look to the revealed trends in market behaviour though this is not always feasible. As a result, a significant literature has been developed around survey methods for estimating individuals' willingness-to-pay (WTP) in the absence of revealed market variation. These methods are now widely used for both developing optimal pricing strategies and also in the forecasting of responses to price changes and for modelling demand functions. This chapter reviews the methods for estimating price sensitivity and willingness-to-pay and discusses some of their relative strengths and weaknesses.

At the highest level, the literature classifies the different methods for estimating WTP into revealed and stated preference methods (see Figure 1). Depending on the type of goods or services in question (and the time and research resources available) both methods can be useful. Revealed preference methods (RP) refer to the observation of preferences revealed by actual market behaviour and represents real-world evidence on the choices that individuals exercise. Moreover, RP data provides valuable information for modelling choice behaviour as the choices reflect decisions that have actually been made.

In some cases, however, the behaviour that is of interest to the analyst may not be observable or currently available. For example, there may be a requirement to evaluate the impact of the introduction of competition where competition has previously not existed or there may be a requirement to quantify the impact of a reduction in competition caused by a proposed merger. In each of these cases it is necessary to make judgements about potential impacts in the absence of real-world evidence on how individual consumers may respond. Stated preference (SP) methods allow examination of such hypothetical situations, which are generated by some systematic and planned design process (Louviere et al., 2000).

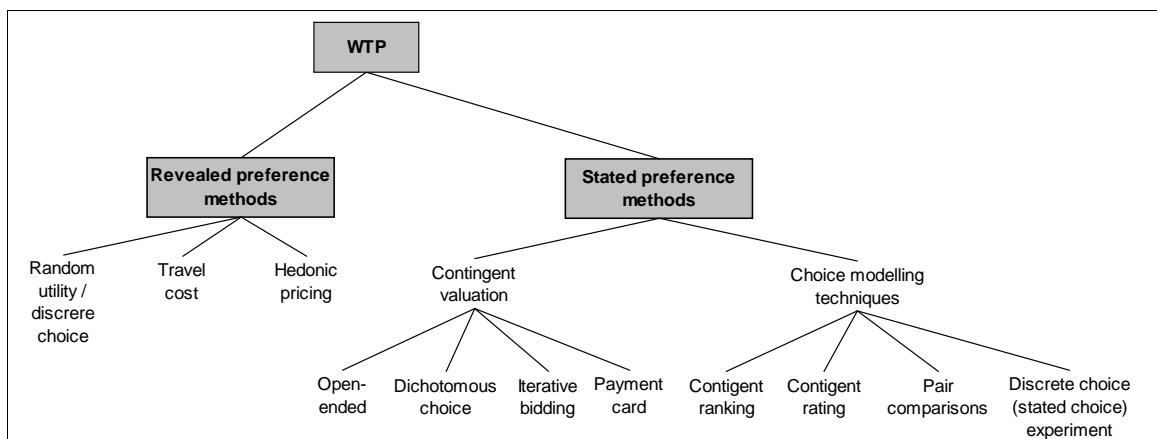


Figure 1 Willingness-to-pay methods (sources: Bateman et al. (2002), Kjaer (2005))

Some key differentiating characteristics of the two methods are shown in Table 1. A prerequisite for using RP methods is that there is an active market for the good or service in question and hence an observable market demand curve (Kjaer, 2005). Also, compared to stated preference (SP) methods that are capable of capturing total economic value<sup>1</sup>, RP methods merely capture “use value” (Kjaer, 2005); ie it is possible within SP methods to also quantify option and non-use value.

**Table 1: Revealed and stated preference methods**

	Valuation method	
	Revealed preference	Stated preference
<b>Approach</b>	Consumers' preferences are revealed through their actions in real markets	Consumers are asked to state their preferences for hypothetical scenarios/alternatives that comprise a set of attributes and different levels of these attributes
<b>Direct methods</b>	Competitive market price (observation of market prices)	Contingent valuation (directly asking individuals their WTP)
<b>Indirect methods</b>	<ul style="list-style-type: none"> <li>• Travel cost method</li> <li>• Hedonic pricing method</li> <li>• Discrete choice</li> </ul>	Discrete choice experiment (estimation of the WTP by use of price variable)
<b>Applicable goods</b>	Real goods	Hypothetical and real goods
<b>Disadvantages</b>	<ul style="list-style-type: none"> <li>• Limited to supply of information regarding values that have been experienced</li> <li>• Limited number of cases where non-market values/goods exhibit a quantifiable relationship with market goods</li> <li>• Choice sets, attributes of choice options and individual characteristics are not controlled and/ designed <i>a priori</i> but rather occur/co-occur</li> </ul>	<ul style="list-style-type: none"> <li>• Observed preferences may not reflect actual behaviour</li> <li>• Absence of incentive for the respondent to provide accurate responses</li> <li>• Incentive for respondent to behave strategically</li> <li>• Overall costly evaluation (more complicated to design and analyse, and also more costly to undertake survey as show material often required for more complex choice task)</li> <li>• Vulnerable to violation of economic decision-making</li> </ul>
<b>Advantages</b>	<ul style="list-style-type: none"> <li>• External validity is maximised because the choices observed are real market choices in which consumers have committed money, time and/or other resources</li> <li>• Low-cost evaluation</li> </ul>	<ul style="list-style-type: none"> <li>• Provides preferences and information that are otherwise impossible to reveal when actual choice behaviour is restricted in some way</li> </ul>

Sources: adapted from Kjaer (2005)

A common issue with RP data is the high degree of collinearity among attributes in market data, making it difficult or impossible to predict the effect of independent variation in an attribute (Kroes and Sheldon, 1988). Using SP methods, it is possible to

<sup>1</sup> The economic value of a product or service is the extent to which people would be prepared to sacrifice something else in order to obtain a quantity of it. Total economic value comprises the sum of use and non-use values. Use values may be direct (consuming the good) or indirect (securing some benefit from the good). In addition to the current use values, individuals may be willing to pay to conserve the option for future use. If the option is related to their own use this WTP reflects option value. Non-use values are known as passive use values and arise in contexts where an individual is willing to pay for a good even though she/he makes no direct use of it.

overcome multicollinearity issues because the analyst has the flexibility to tailor the survey and construct hypothetical situations in order to elicit the desired information and avoid collinearity among attributes.

On the other hand, SP data is also subject to biases that arise from the hypothetical nature of the SP method. Specifically, respondents have no incentive to make a choice in an SP experiment in the same way as they would do in the real situation (Wardman, 1988). In addition, aspects of market choice context, such as search costs, do not exist or are not part of SP experiments. Moreover, a number of well-designed survey steps should be followed in order to minimise the impact of several experimental design issues.

Finally, there are analytical issues to consider when comparing SP with RP. For example, the experimental alternatives in an SP task are defined by the attributes presented, while in RP data there may be attributes observed (or perceived) by the consumer but unobserved by the analyst, and these will be transferred into choice model error terms (Keane, 1997, Kjaer, 2005).

## 2.2 Revealed Preference Methods

The *hedonic pricing* approach follows Lancaster's theory of characteristics of a good, which regards a good (or service) as a set of attributes and considers the value of a good as a function of each attribute of that good (Lancaster, 1966). The value of an attribute is the implicit price (hedonic price) as it cannot be observed in the real market. However, the price of an attribute can be estimated through analysis of the prices of a good that has different quantities of each attribute in the market. Therefore, hedonic pricing is a method of estimating the implicit price that those within the market place on each attribute. The function used to determine the market price of a good based on these attributes is the *hedonic price function* (Hidona, 2002, Kjaer, 2005).

### Box 1: Hedonic pricing method in practice

The hedonic pricing method looks at price and quality variation across goods to infer implicit value of an environmental resource. Goods are considered as bundles of characteristics one of which is the environmental resource of interest. These goods are sold at different prices. Once all other characteristics have been taken into account the differences in price reflect the market value for the environmental resource.

An example could involve differences in property values because of differences in air quality:

- $H_p = a + bQ + cZ$

With  $b$  being the marginal effect on price of a unit change in air quality  $Q$ , and  $c$  the marginal effect of other features ( $Z$ ) that affect house's price such as size, relative location to public transportation, bathrooms, etc.

For example, it may be possible to estimate the value placed on differing levels of noise pollution by examining how house prices vary with proximity to airports once all other aspects that may influence house prices have also been taken into consideration. The main disadvantage with hedonic pricing is multicollinearity between prices and explanatory variables eg poorer quality houses tend to be built in areas subject to noise pollution, along rail lines, etc. Finally, the key prerequisite for such analysis is to have a functioning market in which it is believed that prices are truly set by the market.

The *travel cost* (TC) method is a non-market valuation technique that was developed for use in environmental valuation and has been applied to value recreational demand (eg fishing, hunting, boating and forest visits) (Bateman et al., 2002). Similar to hedonic pricing, the travel cost method seeks to place a value on non-market goods by using consumption behaviour in the market, where travel cost is a measure of the preferences for the good. This is typically applied in situations where there is a desire to estimate a willingness-to-pay value for a service or good which is available at a constant price.

#### **Box 2: Travel cost method in practice**

The travel cost method is mainly used for the valuation of environmental services from recreational sites (eg national parks). This method measures the benefit (WTP) for a recreational experience by examining household expenditures on the cost of travel to a desired recreational site. There are single-site models and multiple-site models. The multiple-site model explicitly considers the fact that people can make trips to alternative recreational sites. This is important because the existence of relevant substitutes will influence the valuations. Data is needed on the characteristics of individuals, the number of visits to the site, information about travel costs etc.

The procedure for conducting a travel cost method analysis involves the following steps:

- Define the benefit to be valued
- Collect the necessary data
- Define the zones of recreation origin
- Calculate the visit costs for each zone
- Determine the visit rate for each zone
- Recognise the model assumptions and constraints
- Apply appropriate statistical methods to calculate WTP

The recreation demand model examines the following relationship:

- $\text{COST OF VISIT} = \text{TRAVEL COST} + \text{OPPORTUNITY COST} + \text{COSTS DUE TO THE DURATION OF THE VISIT}$

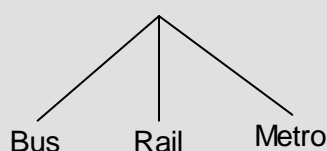
Within the TC approach the individuals accessing the service are segmented into different zones and the travel distance and hence travel cost for each zone is estimated. From this data it is possible to estimate the demand curve to show how demand varies with the cost of access, and by extension how demand may change should the currently fixed price increase or decrease. The rationale behind the TC method is that as the price of access (ie cost of travel) increases the number of visits tend to fall (Kjaer, 2005). However, this method only captures the use value and not the option or non-use values. As with the other RP methods, the TC method is based on actual behaviour, is relatively inexpensive to apply, and results are easy to interpret given that relevant data is available.

Another approach for modelling consumer preferences and indirectly estimating WTP is through discrete choice analysis. Random utility based *discrete choice models* have been developed in this context to both predict individuals' behaviour and to draw inferences about welfare change on the bases of observed choices (McFadden, 1973, Ben-Akiva and Lerman, 1985, Freeman, 2003). The choice underlying the random utility model is the actual choice of the good or service under investigation. It is assumed that individuals make trade-offs among costs and good or service characteristics. Analysis of these choices allow the estimation of the benefits of the characteristics of the good or service.

### Box 3: Discrete choice modelling approach in practice

Discrete choice analysis assumes that individuals consider a number of alternatives and each one selects the outcome with the attribute combination he/she most prefers. By making such choices respondents make implicit trade-offs among different attributes of alternative options. Given a number of responses, a statistical discrete choice model (eg logit) can be used to estimate the relative weight of each attribute. Weights can be monetised by dividing by the marginal utility of cost yielding a WTP or Willingness to Accept (WTA) for changes in an attribute level. Although in many cases WTP is the more appropriate instrument, there are cases where WTA can provide a useful alternative way of framing the valuation question (eg removal of a public service which is currently funded through general taxation, where asking respondents to state a value they would pay to retain it may result in high levels of protest compared to asking how much they would need to be compensated). In either case, policy makers may use these values to understand how alternative scenarios affect the welfare of the sample population.

An example of the discrete choice modelling approach is the choice between different modes of public transport (eg bus, rail and metro). In this example, it is assumed that bus, rail and metro are all possible alternatives available to all individuals.



Model estimation is based on choices made by a sample of public transport users and explanatory variables can be related to the characteristics of the trip maker (eg gender, age, etc), characteristics of the journey (eg, fares, travel times), and those of the transport facility (eg comfort, convenience, safety, etc). If travel cost is included as an attribute, then the ratio of other attributes and cost provide an indirect estimate of WTP, for example willingness to pay to reduce travel times. It is also possible to develop scenarios that involve changes in service attributes eg fares, journey times, service frequency to estimate market shares and consumer surplus.

To date, the Competition Commission's willingness to pay studies have been largely based on stated preference data, mainly because revealed preference (actual choices) data are unavailable or markets do not currently exist. Therefore, the remainder of this report focuses on discussing both the state-of-practice and state-of-the-art of stated preference approaches.

## 2.3 Stated Preference Methods

The family of stated preference methods (see Table 1) can measure the total economic value; that is, SP methods incorporate both non-use value and option value. This characteristic has far-reaching potential as it implies that SP can be used to value potential future or hypothetical (but realistic) goods and interventions (Kjaer, 2005). When considering stated preference methods, the main categories are the Contingent Valuation Methods (CVM) and choice modelling techniques including discrete choice experiments (DCE).

CVM was developed in the US in the 1960s and has found wide applicability in the field of environmental economics. CVM requires a large representative sample of participants and attempts to value a good, service or policy element as a whole. The sum of these values represents a 'social preference function' against which the possible outcomes of the proposal can be weighted (Bruce, 2006). The full array of CVM methods is presented in Box 4. Early CVM studies used single open-ended questions, whereas a framework for dichotomous choice CVM was developed during the 1980s

(eg Hanemann, 1984). The CVM attempts to measure the value of a good as a whole (ie valuing the good in its entirety) by asking people directly their willingness to pay (or willingness to accept).

However, CVM provides no information about the value of the different attributes that comprise the good. This is an important distinction between CVM and choice modelling techniques – the latter estimates the values of the attributes that make up the good – and is one of the main reasons why many analysts consider choice modelling techniques, and DCE in particular, to be superior to CVM (see Table 2).

Moreover, the main criticisms of the dichotomous approach stress the large sample size requirements (and, therefore, high costs) due to the small amount of information collected from each respondent. Also, there are limitations in analysing goods with multidimensional changes (eg price, quality of service and time consideration). Therefore, the shift towards DCE approaches seems pragmatic especially as dichotomous choice and DCE share a common theoretical foundation of Random Utility Theory (RUT) (Kjaer, 2005). However, it should be acknowledged that CVM approaches are still widely used and often provide an attractive method for collecting willingness to pay information in situations where interview duration or difficult fieldwork conditions are a consideration or when it is difficult to develop choice scenarios of the service or policy under consideration.

#### **Box 4: Alternative approaches in CVM**

- *Open-ended (OE)*: "How much are you willing to pay?"
- *Dichotomous choice (DC)*: "Would you pay £X to improve...?"
- *Double-bounded dichotomous choice (DBDC)*: The dichotomous choice question is followed up by another dichotomous choice question depending on the previous answer. This format has been more efficient than its counterpart as more information is collected from each respondent.
- *Iterative bidding (IB)*: Series of dichotomous (yes/no) choice questions followed by a final open-ended WTP question. The bidding increases until the respondent says no.
- *Payment card (PC)*: Respondents select their maximum WTP amount from a list of possible sums presented on a card.

Sources: Bateman et al. (2005), Kjaer (2005)

Choice modelling approaches provide a more direct route to the valuation of the *characteristics* or *attributes* of a good, and of marginal changes in these characteristics, rather than the value of the good as a whole. This may be of particular importance in management decisions, project appraisals and policy appraisals where the decision is sometimes based on changes in the levels that these attributes take. For example, changes in the availability of services may require users to travel farther and use another service with different level of quality and price characteristics.

Early stated preference choice modelling techniques became known in the marketing field as *Conjoint Analysis* (CA)<sup>2</sup> (Green and Srinivasan, 1978). Louviere and Lancsar (2009) argue that "...the [aforementioned] description is so general that it renders the term *conjoint analysis* essentially meaningless as a specific descriptor of any particular SP elicitation approach." Further developments have been influenced by RUT and discrete choice models (McFadden, 1973, Ben-Akiva and Lerman, 1985). RUT is the theoretical foundation of contemporary discrete choice models and became the benchmark for the use of choice techniques in economic literature as it provided the

---

<sup>2</sup> The term originates from the combination of terms "considered" and "jointly"

necessary link between observed consumer behaviour and economic theory (Kjaer, 2005). To distinguish that the latest choice approaches are based on economic theory, the term *Conjoint Analysis* (CA) is no longer used in the economic literature (Louviere and Lancsar, 2009).

Table 2 summarises the key attributes of CVM and choice modelling approaches. Louviere (2000) provides a thorough discussion on the theoretical similarities and differences between conjoint analysis and discrete choice experiments.

**Table 2: Key attributes of contingent valuation and discrete choice experiments**

	Contingent Valuation Methods	Discrete Choice Experiments
<b>Literature</b>	Larger	Emerging
<b>WTP question</b>	Direct question of "What are you willing to pay?" may present cognitive problems	DCE does not directly ask for monetary valuations, but rather asks respondents to choose between alternatives, so it is argued that DCE is easier for people to understand
<b>WTP</b>	Total WTP for the good or service	For individual attributes Relative WTP values for different attributes of a good
<b>Response efficiency</b>	Respondent provides a single response	Each respondent may provide multiple-responses for estimating WTP

Source: Bateman et al. (2002)

Finally, an empirical comparison between Contingent Valuation and Choice Experiments conducted by Adamowicz et al (1998a) investigating the value associated with threatened species found that choice experiments provided a richer description of the attribute trade-offs that individuals are willing to make..

## Choice Modelling Techniques

Within the choice modelling family there are a number of different techniques. In principle, all choice modelling techniques assume that goods or services can be described in terms of their attributes or characteristics and the levels that these take. The focus is on the value placed on these attributes. In stated choice experiments, respondents evaluate and decide which mutually exclusive and multi-attribute alternative they prefer. Each alternative is described by a number of attributes which are offered at different levels across a series of options. For example, the service provided by a water company can be described in terms of quality, taste, odour, annual price, etc. Similarly, a rail service can be described in terms of its cost, timing, and comfort. Choice modelling techniques can be classified into four categories, which reflect differences in theoretical assumptions, methods of analysis and experimental design procedures (Adamowicz and Boxall, 2001, Bateman et al., 2002, Kjaer, 2005):

- Discrete choice or stated choice experiments,
- Contingent ranking,
- Contingent rating, and
- Paired comparisons.

Each of these is described below.

In *discrete choice experiments* (DCEs), respondents choose one alternative out of two or more alternatives on offer. Each respondent may be asked to repeat the choice exercise multiple times; with the levels of the attributes changing according to an experimental design (for more details see Section 3.3). In some cases, each choice set may also include a base case or *status quo* alternative, so the respondent has the choice of remaining with the current service or product. In other cases, where the policy being examined suggests that it may be impossible to stay with the *status quo* (eg a merger that would lead to a closure of local facilities or extensive changes in pricing) it is necessary to consider a realistic opt-out option within the set of alternatives which can be chosen if none of the other alternatives offered are acceptable eg “I would no longer purchase this service”.

Participants in discrete choice experiments are asked to make trade-offs between changes in the levels of a range of attributes which, in the case of a WTP study will typically include a monetary cost attribute (see Figure 2). From this trade-off data the analyst can gain insight into their willingness to pay (or accept). In terms of the theoretical basis, DCEs only collect information about the alternative chosen, and not the full ranking of alternatives. Therefore, the data are considered weakly ordered. Also, the econometric technique used in DCE can be represented in a way that is exactly parallel to economic theory of rational, probabilistic choice (Bateman et al., 2002). Finally, estimates of compensating and equivalent surplus can be derived from the “output” of the technique.

If the following public transport options were available, which would you choose for your journey between Cambridge and Manchester?				
	Coach	Rail	Air	
Expected total travel time	7 hrs 55 mins	3 hrs 50 mins	2 hrs	I would not make the journey
Waiting time	10 min	5 min	1.5 hrs	
Ticket price (one-way trip)	£40.20	£81.60	£70.00	
Service frequency	Once a day	Every 30 mins	Twice a day	
Interchanges	0	2	0	
I would prefer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 2: An example of a discrete choice experiment question

By contrast, in a *contingent ranking* exercise respondents must rank all of the alternative options on offer and, in this case, the data are said to be strongly ordered. (see Figure 3). While contingent ranking data offers much richer information than discrete choice experiments, they are more cognitively demanding (Louviere et al., 2000). Contingent ranking can also be approached as a sequential choice process where respondents are first asked to choose the most preferred alternative out of a total of N alternatives. Next, the chosen alternative is removed from the choice set and respondents are asked to choose the most preferred option out of the remaining N-1 options. In order to interpret results from contingent ranking experiments in standard welfare economic terms, a *status quo* or ‘choose none’ option must be included in the choice set. Otherwise, respondents are forced to choose one of the alternatives on offer, which they may not prefer at all.

If the following public transport options were available, could you rank them according to your preferences, assigning 1 to the most preferred, 2 to the second most preferred and 3 to the least preferred or your journey between Cambridge and Manchester?

	Coach	Rail	Air
Expected total travel time	7 hrs 55 mins	3 hrs 50 mins	2 hrs
Waiting time	10 min	5 min	1.5 hrs
Ticket price (one-way trip)	£40.20	£81.60	£70.00
Service frequency	Once a day	Every 30 mins	Twice a day
Interchanges	0	2	0

Your ranking:

1 \_\_\_\_\_

2 \_\_\_\_\_

3 \_\_\_\_\_

Figure 3: An example of a contingent ranking question

In a *contingent rating* experiment, respondents are presented with one alternative at a time and are asked to rate each one separately on a semantic or numeric scale (eg low preference - high preference, 1 - 10, see Figure 4). Each respondent may provide ratings to a series of alternative choices in which attributes again vary. The degree of task complexity in contingent rating is even higher than contingent ranking or discrete choice experiments as respondents have to place a value (also called strength of preference) on each alternative (Louviere et al., 2000). Compared to contingent ranking and discrete choice experiments, contingent rating does not involve direct comparisons between alternatives; however, it allows respondents to score alternatives equally and, thus, indicate indifference between alternatives.

1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Very low preference					Very high preference				
<p><b>On the scale below, please show how strongly you would prefer the following policy option</b></p>									
					<b>Coach</b>				
Expected total travel time					7 hrs 55 mins				
Waiting time					10 min				
Ticket price (one-way trip)					£40.20				
Service frequency					Once a day				
Interchanges					0				

Figure 4: An example of a contingent rating question

Finally, *pairwise comparison* exercises ask respondents to choose their preferred alternative from a set of two choices and to indicate the strength of their preference in a numeric or semantic scale (see Figure 5). This approach can be seen as a combination of a discrete choice experiment and rating exercise where a respondent chooses the most

preferred alternative, but also rates his/her strength of preference. However, frequently no 'choose none' option is included in these exercises and this leads to a task that can provide preference information but cannot be considered within a wider welfare economics framework.

**Which of these two options would you prefer for your journey between Cambridge and Manchester?**

	<b>Coach</b>	<b>Rail</b>
<b>Expected total travel time</b>	7 hrs 55 mins	3 hrs 50 mins
<b>Waiting time</b>	10 min	5 min
<b>Ticket price (one-way trip)</b>	£40.20	£81.60
<b>Service frequency</b>	Once a day	Every 30 mins
<b>Interchanges</b>	0	2

**I would prefer:**

*Definitely Coach*     

*Possibly Coach*     

*Cannot choose*     

*Possibly Rail*     

*Definitely Rail*     

*Choose none*     

**Figure 5: An example of a pairwise comparison question**

More recently, contingent ranking style experiments have been making a resurgence through the *best-worst scaling* (BWS) approach, which Louviere and colleagues have been advocating as a variant on the conventional stated choice task (Finn and Louviere, 1992, Flynn et al., 2007). In the BWS approach respondents are asked to indicate the two extremes of their preference when considering a single set of attributes ie which is the best and then which is the worst attribute. When more than 4 alternative options are involved the analyst can also ask respondents for their choice of second best and second worst attribute. Therefore, the elicited information is the value placed on individual attribute levels rather than preferences for alternatives made up of bundles of attribute levels.

**Which of the following five travel characteristics for your coach journey between Cambridge and Manchester would you rate as being the best and which as being the worst?**

<b>Best</b>		<b>Worst</b>
	Expected total travel time is 7 hrs 55 mins	
	Waiting time 10 min	
	Ticket price (one-way trip) £40.20	
	Service frequency Once a day	
	Interchanges 0	

**Figure 6: An example of a best-worst scaling question**

In terms of practical application, the academic literature over recent years has largely converged around discrete choice (or stated choice) experiments. A number of authors argue that this provides a task that most closely mirrors that which people face in real life ie a single choice between a set of alternatives. (Louviere et al, 2000, Hensher et al, 2005). Moreover, DCE allow for indirect estimation of WTP, which is the recommended approach by the HM Treasury (Cave et al, 1993) for the valuation of changes in quantity or public services. On the other hand, BWS is a useful method for estimating the values (not preferences) placed on certain attributes on a common interval scale, but it cannot be used to estimate future demand and elasticities.

## 3. DESIGNING STATED PREFERENCE EXPERIMENTS

### 3.1 Introduction

This chapter focuses in more detail on the issues around the design of survey questions in contingent valuation and discrete choice experiments.

### 3.2 Designing Contingent Valuation Questions

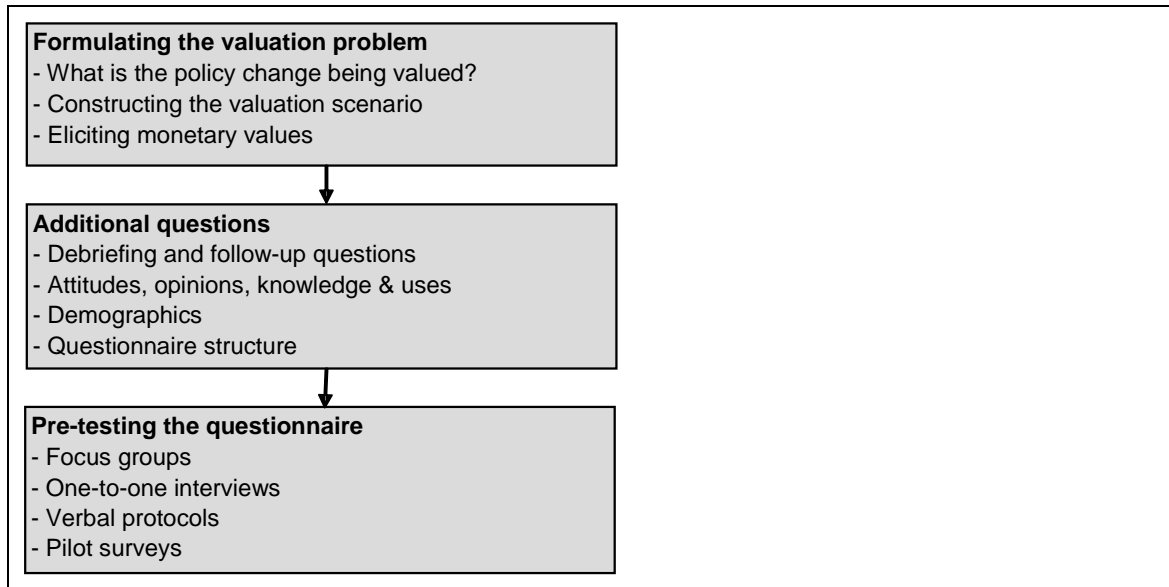
The objective of a contingent valuation (CV) questionnaire is to elicit preferences in monetary terms, more specifically the maximum WTP or minimum Willingness to Accept (WTA)<sup>3</sup> for changes in the quantity or quality of a good or service from a random sample of respondents. These changes may refer to a hypothetical or an actual good or service.

The design of a CV questionnaire requires special attention as some of its features differ from standard polling and marketing research surveys. Bateman et al. (2002) point at three differences. Firstly, CV questionnaires ask respondents to consider how a change in a good or service might affect them. The policy change (real or hypothetical) should be perceived as being realistic and feasible and is described in detail before respondents are asked to evaluate it. Secondly, the hypothetical good or policy may be complex and unfamiliar to respondents which may result in inconsistent and unreliable observations. Finally, respondents are asked to state their WTP or WTA for the change in the good or policy in question.

As shown in Figure 7, the design of a CV questionnaire involves three main stages: formulating the valuation problem, additional questions and pre-testing of the questionnaire.

---

<sup>3</sup> In discrete choice experiments, the following criteria must apply: the number of alternatives in the choice set is finite, the alternatives are mutually exclusive and the set of alternatives is exhaustive (all possible alternatives are included) (see Train, K. (1993) *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, 3rd Edition, MIT Press, London.). In practice, the criteria that the choice set is exhaustive can in many cases be met by including a “none of the above” or a “decline to purchase good/service” alternative within the choice set.



**Figure 7: Design stages in CV questionnaire** (source, Bateman et al., 2002)

The ‘Additional questions’ and ‘Pre-testing the questionnaire’ stages are common with the design of a choice experiment and, thus, are discussed in Appendix B. In this section, the focus is on formulating the valuation problem in a contingent valuation study.

### **What is the policy change being valued?**

This *first stage* involves the identification of the good to be valued, constructing the valuation scenario and eliciting monetary values (see Figure 7). At this stage, the analyst must have a clear understanding of what *policy change is to be valued* and which quality and quantity change of the good or service is of interest. In many cases, formulating the valuation problem involves a number of issues including:

- scientific uncertainty about the problem
- increased complexity and multidimensionality of the change that cannot be adequately described in the context and timeframe of a questionnaire, and
- difficulty in describing the changes in a textual format (eg air pollution, noise, etc).

To control for, or minimise these issues, the analyst should draw from existing information and relevant experts.

### **Constructing the valuation scenario**

Having defined the policy change or change in the good or service of interest, the next step is to *construct the valuation scenario*. This step requires special attention as information about the good or service, the wording and type of the valuation questions, and the payment mechanism are likely to influence the accuracy and reliability of individuals’ responses. In practical terms, a valuation scenario should include descriptions regarding (Bateman et al, 2002):

- The change in the good, service or policy of interest,
- The constructed market (for non-market goods), and
- The method of payment.

Each of these is described below.

- *Change in the good, service or policy of interest*

Describing the change in the good, service or policy of interest entails a number of steps. Firstly, it requires a meaningful and understandable description of the attributes of the good under investigation. The description may be detailed and include all attributes, focused on the most important or familiar valuable attributes of the good, or merely describe the good in general, ignoring its attributes. The objective is to provide the necessary information in a manner that does not result either in a vague description of the good or information overload.

Bateman et al (2002) advise that an accurate scenario can be best specified through using a balanced approach that focuses on the most important or familiar attributes (with short descriptions that capture respondents' attention). Secondly, part of the scenario description should be the *available substitutes* for the good and alternative expenditure possibilities that may affect respondents' values. Thirdly, the scenario should include a description of the proposed policy change and how the different attributes of the good of interest will change accordingly.

In particular, it is important to include a clear description of the *reference* (baseline) and *target* (proposed change) levels of each attribute of interest. Also, a description of the *scope* of the policy change requires particular care as it has been seen as one of the main criticisms of the validity of the CV approach. Therefore, a clear, detailed and meaningful definition of the scope of the proposed policy changes is required. Finally, the descriptions of the good and the policy change can be enhanced by textual information such as images, maps, charts and graphs (Bateman et al., 2002). These media help to draw the respondents' attention and convey complex information in a simpler and more consistent way.

In some cases, the proposed policy may involve *multi-dimensional changes* and not just in the attributes of a single good. Therefore, the analyst needs to find out the total or aggregate value of the policy. The description of the policy scenario in this case should provide accurate information about each and all its relevant components and how these would change the *status quo*. Changes may be *simultaneous* or *sequential*. When the proposed policy concerns simultaneous changes then each individual change of a relevant component should be presented as part of the broader package.

This can be done in two ways: one way is to describe the overall policy with all its components and then ask respondents to provide their valuation of a specific component. Another approach is to ask respondents to value the overall policy and then to partition that total value across its components. The obvious limitation in the latter approach is respondent fatigue, which will limit the number of policy changes and components that can be valued.

In the case that the policy change of interest involves a sequence of changes then its value depends on the point of the sequence in which it appears and may become progressively smaller as it is placed deeper and deeper in the valuation sequence. To obtain valid estimates, respondents should first be aware of all the elements of the sequence so that they can assess the total value of the policy. The sequential approach involves many uncertainties which may result in a very complicated

valuation exercise. The choice modelling approach, described in Section 3.3, presents an effective way of overcoming the issues related to multi-dimensional policy changes and estimating the substitution parameters between goods (Bateman et al, 2002).

- *The constructed market*

The constructed market refers to the social context in which the policy change takes place. A number of elements may be required to describe this context including:

(a) The institution responsible for the change. This can be a government, local council, research institute, business, charity etc. Respondents' WTP is likely to be affected by institutional arrangements as respondents may hold positive or negative views about the institution involved.

(b) Technical and political feasibility of the change is also likely to affect responses. Respondents are more likely to provide reliable valuations if the proposed scenario of policy change is feasible.

(c) Conditions for the provision of the good. This includes respondents' perceived payment obligation and respondents' expectations about provision. The former implies several possibilities:

- respondents may believe they will have to pay the amount they state;
- they may think the amount they have to pay is uncertain, or
- they may be told that they will pay a fixed amount or a proportion of the costs of provision.

With regard to respondents' expectations about provision, these relate to whether they believe (or not) that the provision of the good is conditional on their WTP amount.

(d) Timing of provision, when and how long the good or service will be provided.

(e) Addressing who will pay, benefit and lose from the proposed policy change of interest. Payers can be users, tax payers, donors, industry, etc.

- *The method of payment*

The method of payment includes a number of different aspects:

- *Choice of benefit measure.* There are four types of benefit measures that can be elicited: (a) *WTP to secure a gain*, (b) *WTP to avoid a loss*, (c) *WTA to tolerate loss*, and (d) *WTA to forgo a gain*. The choice of the benefit measure is rather case-specific and depends on the characteristics and context of the change valued. It is worth mentioning that the choice of the measure may impact the resulting valuations. For example, empirical evidence has shown that WTA responses are substantially higher than the WTP responses, even for fairly familiar goods of rather low monetary value (Kahneman et al, 1990, Bateman et al, 1997).

**Table 3: Example questions of welfare change measures** (adapted from Bateman et al. (2002))

Welfare measure	Example question
WTP to secure a gain	Would you be willing to pay £X on top of your annual water bill for an increase in the water quality that comes to your home?
WTP to avoid a loss	Would you be willing to pay £X on top of your annual water bill to avoid a decrease in water quality that comes to your home?
WTA to tolerate loss	Would you be willing to accept £X to accept a decrease in water quality that comes to your home?
WTA to forgo a gain	Would you be willing to accept £X instead of an increase in water quality that comes to your home?

- *Payment vehicle.* Describes the way in which the respondent is (hypothetically) expected to pay for the good. While there are no precise rules for choosing between payment vehicles, the nature of the good being a local good can determine the payment vehicle (eg exclude the case of using national tax as payment vehicle). As shown in Table 4, the payment vehicle can be coercive or voluntary. Using coercive payment vehicles (eg tax scheme) raises issues of accountability, and trust in the government, the potential exclusion of non-tax payers or it may not be credible when the scenario concerns WTA (eg corresponding to tax rebate). On the other hand, with voluntary payments, respondents have an incentive to overstate their WTP to secure provision and to reduce real payments relative to their WTP once the provision is secured (free-riding on the donations of others). The simple guideline is to use the payment vehicle which is likely to be employed in the real world decision context.

**Table 4: Types of payment vehicles** (source, Pearce and Ozdemiroglu, 2002)

Coercive	Voluntary
Tax (national, local)	Donation to trust fund
Fee/charge	Gift
Price increase	

- *Target and reference payment levels.* The reference level is respondents' disposable income, and reflects budget constraints. The target payment level is the respondents' maximum WTP or minimum WTA. If the respondent already pays for the provision of the good then it should be made clear that any new payment will be over and above what is currently paid.
- *Individual or household payment.* In principle, the reference income should be consistent with the unit of analysis, this being either individual or household valuations. In eliciting WTP or WTA of individuals, there is a risk that individuals being interviewed may not have an independent income. On the other hand, household payments may be a problem when respondents are uncertain about the reference household income levels and about other household members' values.
- *Timing of payment.* Finally, payments may be expressed as a one-off lump sum, a yearly, monthly, daily or per visit/use amount, etc. The lump sum payment approach tends to eliminate many of the sequencing problems – the good bought and paid as one off does not enter the sequence of possible new goods. On the other hand, there is a need to describe the timescale over which repeated

payments will take place. An easy approach is to present respondents with monthly or annual amounts in the elicitation question.

### Eliciting monetary values

The elicitation question can be asked in a number of different ways. The choice of elicitation format is of considerable importance as different elicitation formats typically produce different estimates. However, in all approaches respondents must be reminded of substitute goods, the need to trade-off money for benefits, and their budget constraints (Bateman et al., 2002). As shown in Box 4 and the following tables, the most widely used methods include:

- **Open-ended** elicitation asks respondents their maximum WTP (Table 5),
- **Iterative bidding** offers several rounds of discrete choice questions or bids with the final question being an open-ended WTP question (Table 6),
- **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 (Table 7), and
- **Single- and double-bounded dichotomous choice** (Table 8 and Table 9). In the former, respondents say ‘yes’ or ‘no’ to a single WTP amount or bid whereas, in the latter, respondents say ‘yes’ or ‘no’ to a stated sum and are then asked to say ‘yes’ or ‘no’ to higher/lower bids.

**Table 5: Format, advantages and disadvantages of the open-ended elicitation** (*adapted from (Pearce and Ozdemiroglu, 2002)*)

<b>Open-ended elicitation</b>	
What is the maximum amount that you would be prepared to pay on top of your annual water bill to improve the quality of drinking water that comes to your home?	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> <li>▪ Straightforward</li> <li>▪ No anchoring bias; does not provide respondents with cues about what the value of the change might be</li> <li>▪ Very informative as maximum WTP can be identified for each respondent</li> <li>▪ Requires relatively straightforward statistical techniques</li> </ul>	<ul style="list-style-type: none"> <li>▪ It leads to large non-response rates, protest answers, a high proportion of zero answers (which can be protest responses), and outliers (ie unrealistically large bids) and, generally, to unreliable responses. In some cases, respondents may find it difficult to come up with their true maximum WTP for a change they are unfamiliar with and have never thought about valuing before. Most market transactions involve a decision on whether to buy a good or not, rather than stating maximum WTP values.</li> </ul>

**Table 6: Format, advantages and disadvantages of the iterative bidding elicitation** (adapted from (Pearce and Ozdemiroglu, 2002))

<b>Iterative bidding</b>	
<p>Would you pay an additional £5 every year through your annual water bill to improve the quality of drinking water that comes to your home?</p> <p>If Yes: Interviewer keeps increasing the bid until the respondent answers "No". The maximum WTP is elicited.</p> <p>If No: Interviewer keeps decreasing the bid until respondent answers "Yes". The maximum WTP is elicited.</p>	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> <li>▪ May encourage respondents to consider their preferences carefully</li> </ul>	<ul style="list-style-type: none"> <li>▪ Risk of anchoring bias (ie respondent may be influenced by the starting values and succeeding bids used)</li> <li>▪ Leads to large numbers of "yea-saying" bids and outliers</li> <li>▪ Bidding games are not suitable for mail surveys</li> </ul>

**Table 7: Format, advantages and disadvantages of the payment card elicitation** (adapted from (Pearce and Ozdemiroglu, 2002))

<b>Payment card</b>	
<p>Which of the amounts listed below best describes your maximum willingness to pay on top of your annual water bill to improve the quality of drinking water that comes to your home?</p>	
<p>£0 £0.50 £1 £2 £3 £4 £5 £7.50 £10 £12.50 £15 £20 £30 £40 &gt;£40</p>	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> <li>▪ Provides a context to the bids, while avoiding starting point bias at the same time</li> <li>▪ The number of outliers is reduced compared to open ended elicitation and iterative bidding</li> <li>▪ Some versions of the payment card show how the values on the card relate to actual household expenditures taxes (benchmarks)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Vulnerable to biases relating to the range of the numbers used in the card and the location of the benchmarks</li> <li>▪ Cannot be used in telephone interviews, but feasible to implement within a phone post phone survey</li> </ul>

**Table 8: Format, advantages and disadvantages of the single-bounded dichotomous choice** (adapted from (Pearce and Ozdemiroglu, 2002))

<b>Single-bounded dichotomous choice</b>	
<p>Would you pay £5 every year on top of your annual water bill to improve the drinking water quality that comes to your home? (The price is varied randomly across the sample). Yes / No</p>	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> <li>▪ This task provides less cognitive burden for the respondents who only have to state "Yes" or "No" to a given price. The approach mimics the way a respondent would decide whether or not to buy a good at a certain price. At the same time it provides a truthful revelation of preferences under certain circumstances (incentive compatibility). That is, it is in the respondent's strategic interest to accept the bid if their WTP is greater or equal than the price asked and to reject otherwise ie ensuring that the respondent tells the truth.</li> <li>▪ Minimises non-response and avoids outliers</li> <li>▪ The approach has received endorsement of the National Oceanic and Atmospheric Administration (NOAA) panel (Arrow et al., 1993)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Empirical studies have shown that WTP values obtained through dichotomous choice elicitation are significantly larger than those resulting from comparable open-ended questions</li> <li>▪ Some degree of "yea-saying" is possible</li> <li>▪ Dichotomous choice formats are relatively inefficient in that less information is available from each respondent (the analyst 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</li> <li>▪ There may also be starting point bias</li> </ul>

**Table 9: Format, advantages and disadvantages of the double-bounded dichotomous choice** (adapted from (Pearce and Ozdemiroglu, 2002))

<b>Double-bounded dichotomous choice</b>	
<p>Would you pay £5 every year on top of your annual water bill to improve the drinking water quality that comes to your home? (The price is varied randomly across the sample). If Yes: And would you pay £10? If No: And would you pay £1?</p>	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> <li>▪ 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 they accepted to pay £5 in the first question but rejected £10 in the second.</li> </ul>	<ul style="list-style-type: none"> <li>▪ 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 subsequent dichotomous questions (seeking to provide the double-bound) may not be viewed by respondents as being exogenous to the choice situation, and the added possibility of anchoring and "yea-saying" biases.</li> </ul>

Overall, payment cards (Table 7) and dichotomous choice formats (Table 8 and Table 9) are the most recommended CV elicitation methods (Bateman et al., 2002, Pearce and Ozdemiroglu, 2002). Payments cards offer more information less cognitive burden to respondents and are cheaper to implement than dichotomous choice. In addition they are superior to both direct open-ended questions and bidding games. Dichotomous choice formats may be incentive compatible and facilitate respondents' valuation task.

### 3.3 Designing Choice Modelling Experiments

As was shown in Chapter 2, the various categories in choice modelling approaches exhibit a number of fundamental differences in both the theoretical assumptions and the structure of the choice tasks. However, a number of common assumptions and design stages do exist. In particular, all choice modelling experiments are based on data generated through a systematic and planned experimental design process in which the attributes and their levels are pre-defined and varied to create preference or choice alternatives (Louviere et al., 2000).

Figure 8 shows the design steps used to generate a choice modelling experiment. The next few sections briefly describe the main objectives and considerations of each step.

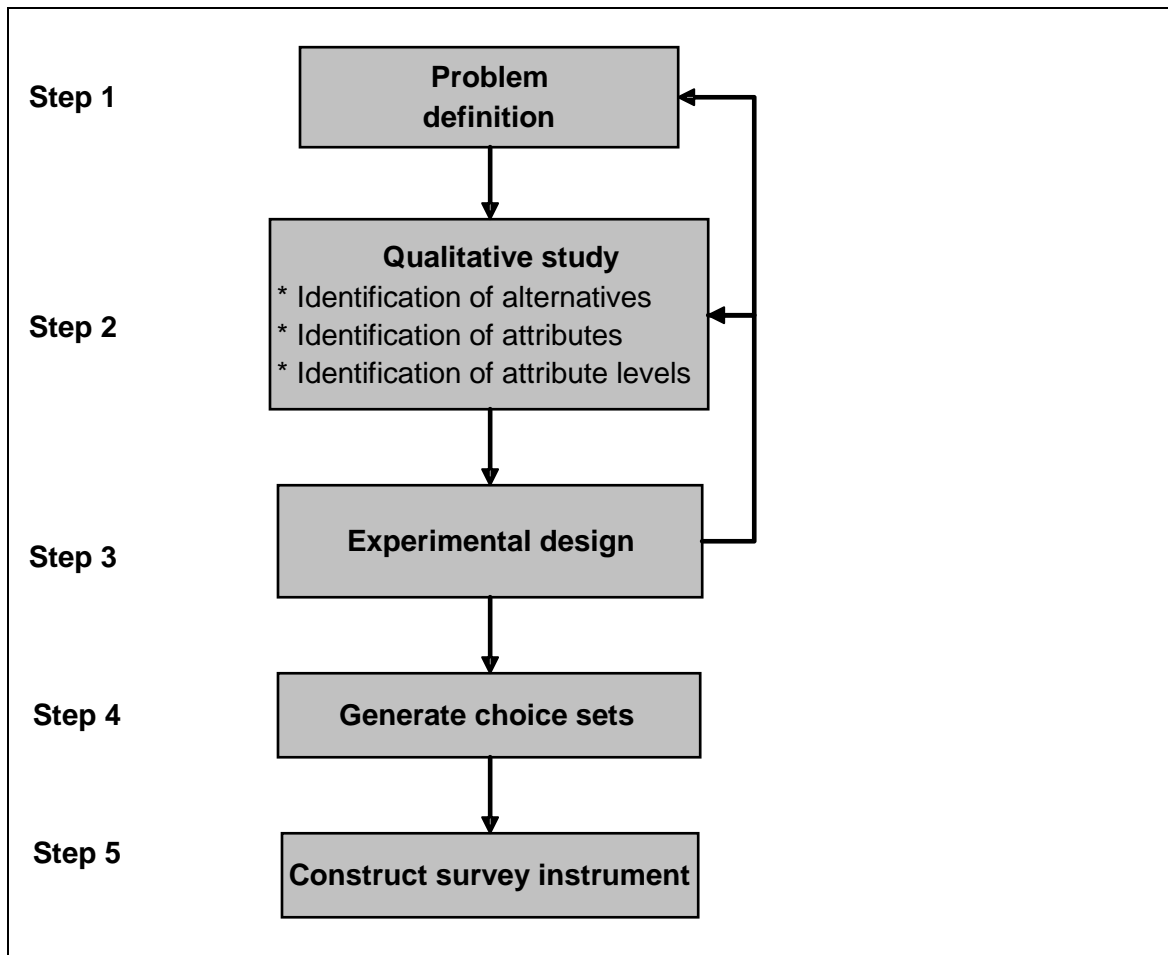


Figure 8: Design process in choice modelling experiments (source: Hensher et al. (2005))

#### Step 1: Problem definition

The design process begins with problem definition, which allows the analyst to define the research question and set the objectives and outcomes of the research project. A better understanding of the problem may be achieved through the following questions (Louviere et al., 2000, Hensher et al., 2005):

- What are the existing alternatives (goods or services) available?
- What are the attributes of existing alternatives?

- What are the determinants (most important factors) influencing demand for existing alternatives?
- Is the choice process consistent over time or is it seasonal?
- What is the population under study?
- What is the output required? (ie valuation or demand modelling study).

In this context, an “alternative” is a good or service that the respondent could have chosen in the choice context of interest, be that the one that they are known to have selected in an existing choice situation (observed through revealed preference data) or the other options which they could have considered instead in making their choice.

The aforementioned questions would determine the particular methodological approach to be employed. In particular, sufficient understanding of the problem and research question(s) would allow the analyst to decide whether a stated preference approach is necessary and, therefore, proceed to the next step of the choice modelling design.

## Step 2: Supporting qualitative study

The objective of this step is to:

- Define the list of potential alternatives<sup>4</sup>,
- Define the list of attributes, which must be mutually exclusive,
- Define attribute levels.

Each of these is described below:

- *List of Alternatives*

As shown in Table 10, the *number of alternatives* required in contingent ranking and paired comparisons are fixed. On the other hand, contingent ranking and discrete choice experiments frequently need to draw on qualitative research in order to identify each and every possible alternative that may exist. Identification of all possible alternatives is particularly important in the discrete choice experiment exercises in order to meet the global utility maximising criteria of Random Utility Theory. It should be noted that failure to identify all alternatives produces a constraint, a threshold on utility maximisation outcome (Louviere et al., 2000). However, in some cases, it might be necessary to select (eg exclude ‘insignificant’ alternatives), merge some alternatives or assign a randomly sampled number of alternatives to respondents in order to reach a manageable number of alternatives to study (Hensher et al., 2005). Another approach is to develop experiments without naming the alternatives (unlabelled alternatives) (Hensher et al., 2005); within such an approach respondents are asked to consider two or more alternatives which differ in terms of the levels at which the attributes are presented but which are not given any additional ‘branding’; such experiments may still require a ‘choose neither’ option to be exhaustive.

---

<sup>4</sup> In discrete choice experiments, the following criteria must apply: the number of alternatives in the choice set is finite, the alternatives are mutually exclusive and the set of alternatives is exhaustive (all possible alternatives are included) (see Train, K. (1993) *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, 3rd Edition, MIT Press, London.)

In identifying the relevant alternatives it may be useful to consider the evidence that may be obtained from a range of different sources, including: secondary data sources, literature reviews, in-depth interviews and focus groups (Louviere et al., 2000, Amaya-Amaya et al., 2008).

**Table 10: Number of alternatives in choice modelling approaches**

Choice modelling approach	Number of alternatives per exercise
Discrete choice or stated choice experiments	2 or more (may include a status quo)
Contingent ranking	2 or more (may include a status quo option)
Contingent rating	1
Paired comparisons	2

Source: Bateman et al. (2002)

- *List of Attributes*

Having identified the number of alternatives, the next step is to determine the *list of attributes and attribute levels* to describe the alternatives. While a standard method for selecting attributes does not exist, the following two criteria may guide their selection (Bennett and Blamey, 2001):

- Attributes should be relevant to policy makers, and
- Attributes should be meaningful and important to respondents.

There are no strict rules with regard to the number of attributes to be included; some researchers believe that all relevant attributes should be included whereas others argue that the maximum number of attributes in a single DCE should be no more than eight (Ortuzar and Willumsen, 2001). In particular, the objective is to include a smaller number of attributes because the required sample size increases exponentially with the number of attributes (Bateman et al, 2002), whilst balancing against this the concern that the exclusion of relevant attributes may result in biased estimates and inaccurate welfare measures (Kjaer, 2005).

Deciding on the list of attributes requires particular care and can benefit from literature reviews, focus group discussions or direct questioning (Bateman et al, 2002). It may be necessary to consider a mix of both common (generic across all alternatives) and alternative-specific attributes to describe the various choice alternatives if there are factors which are important for some of the alternatives. Even for common attributes there may be challenges in defining the attributes as the appropriate values may significantly differ from alternative to alternative depending on the context being considered. For example, the travel cost may appear reasonably similar for a train and a car journey for a single traveller but may in fact be significantly different when the costs for the entire group travelling is taken into account; such issues need careful consideration in the design of the task to ensure that it appropriately represents the factors which the respondent would need to consider when making their choices. The inclusion of a monetary cost attribute in choice experiments allows the estimation of WTP. However, in doing so the values of price need to be credible and realistic so as to minimise the possibility for strategic behaviour (Bateman et al, 2002).

The definition of some attributes may involve ambiguities in their meaning and measurement between the analyst and the respondents; for example, the definition

of ‘comfort’ when choosing mode of transport. It is, therefore, important for the analyst to consider approaches for reducing any ambiguity in definitions; for example the ‘comfort’ example mentioned previously may warrant the use of some photographic show material to explain different seating configurations prior to, or as part of, the choice task to provide greater consistency in the understanding of the various levels of comfort being discussed.

A related consideration is *inter-attribute-dependence*<sup>5</sup>, which refers to the cognitive perceptions decision makers attach to the attribute descriptions provided. An example of this is the *price-quality* interaction where respondents may perceive that higher-priced alternatives are of higher quality. While it is possible to generate an experimental design that allows estimation of the importance of price on choice independently from quality, respondents may not treat these two attributes independently (Hensher et al, 2005). Hensher et al (2005) propose a number of strategies to overcome the issue of inter-attribute-dependence including designs of nested attributes. However, a simpler approach would be to include attributes that may act as proxies for other attributes and use the most appropriate ones for the study (eg include crowding to describe comfort in a choice situation that involves a rail alternative).

- *Attribute levels*

The next task is to decide on the number of *attribute levels* and assign *attribute-level labels* to each of the attribute levels selected for inclusion in the experiment. Attribute levels are the levels assigned by the analyst to an attribute at the experimental design stage and have no particular meaning to the respondent (Hensher et al, 2005). On the other hand, attribute-level labels are the narratives assigned to each attribute level for a given attribute. As shown in Table 11, attribute-level labels can be either qualitative or quantitative. Quantitative attribute-levels can be represented in either absolute or relative terms; for example, 20% more than the current price of the product (Bennett and Blamey, 2001). Ryan (1999) provides three considerations when choosing attribute-level labels:

- Attribute-level labels should be *plausible* to the respondents,
- Attribute-level labels should *provide meaningful information (be actionable) and capable of being traded*,
- The labels must be constructed so that respondents are willing to *trade-off* between combinations of the attributes.

**Table 11: Attribute levels and quantitative and qualitative attribute-level labels**

Attribute Level	Quantitative	Qualitative
	Attribute: PriceLabel	Attribute: ColourLabel
1	£20	"Green"
2	£40	"Yellow"
3	£60	"Red"
4	£80	"Brown"

The identification of attribute levels and attribute-level labels requires a number of decisions including (Hensher et al, 2005):

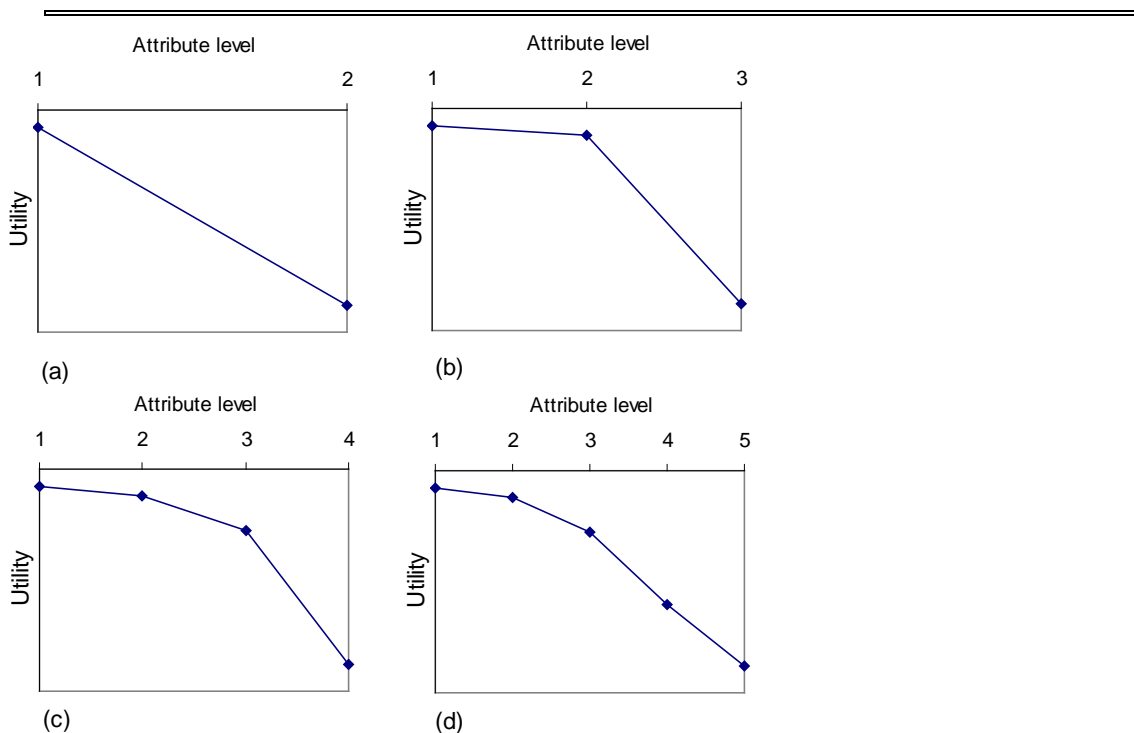
- How *many attribute levels* to include per attribute,

<sup>5</sup> Also referred to as inter-attribute correlation (Hensher et al, 2005)

- How to define *level range* (ie the interval between levels),
- How to identify *extreme ranges* (ie the end-points of the levels).

The number of attribute levels need not be the same across all attributes (Amaya-Amaya et al, 2008). In principle, more attribute levels are more likely to provide richer and more accurate information. When, for example, an attribute has only two levels then the marginal or part-worth<sup>6</sup> utility function can only be linear. For example, Figure 9a shows the marginal (dis)utility of price when the attribute has two levels<sup>7</sup>. Figure 9b, c and d illustrate how including more levels enables the analyst to identify non-linear utility relationships (Hensher et al, 2005, Amaya-Amaya et al, 2008).

Often the analyst has to compromise in terms of the number of attribute levels. More levels provide a better understanding of the relationship between an attribute's level and the marginal utility, but also have a direct impact on the complexity and size of the required experimental design, which in turn places demands on the sample size required to estimate robust models. As is the case with the number of attributes and alternatives, the size of the experimental design increases exponentially with the number of attribute-levels. The general practice of using 3 to 4 levels would provide the analyst with a good understanding of the underlying relationship and thus detect if any non-linear relationships exist. However, in cases where there is particular interest in establishing the potential for non-linearities over a wide range of, for example, price increases, it may be desirable to include a greater number of attribute levels.



**Figure 9: An example of marginal (part worth) utilities of price**

<sup>6</sup> Refers to the level of utility derived from a single attribute at varying levels

<sup>7</sup> In this example, it is assumed that price-levels are in order with price-level labels and in the following sequence: price-level 1 < price-level 2 < price-level 3 < price-level 4 < price-level 5. In general, actual coding of price-level labels need not necessarily follow the order of price levels.

Level range<sup>8</sup> or the interval between attribute-level labels is also an important issue to address as an improper attribute-level range might result in biased results. When the level-distance is too narrow respondents may perceive the difference to be insignificant, whereas if it is too wide respondents may perceive the difference to be highly significant. The former may lead to increased noise in the data and poor estimates of the differences, whereas the latter may lead to non-trading behaviour and can have a significant impact on model results (Hess et al, 2008). Here, it is important to note that an insignificant coefficient does not necessarily mean that the attribute is unimportant to respondents, but it is not possible to demonstrate a significant relationship between the given level and the choice (Kjaer, 2005).

A related issue is that of how to identify *extreme ranges* (minimum and maximum levels) of the attribute levels. Information on extreme ranges may be provided through qualitative research. Hensher et al (2005) suggest that defining attribute levels beyond the identified range is feasible but should be done with care as there is always the risk of misinterpretation and disbelief by the respondents. This has important implications for the price attribute as choosing an inappropriate cost interval may result in under- or over-estimated WTP values. Alternatives with prices being too low are more likely to be chosen resulting in a very high price coefficient and, hence, deflated monetary values for attribute levels. On the other hand, prices being too high will almost always be rejected and estimated coefficients will be small or zero (Bateman et al, 2002). However, in order to understand the full distribution of WTP (and the population/sample mean) it is necessary to include attribute-level values sufficient to stimulate trading across the whole population.

The above discussion has been heavily based on the presumption that attributes and, subsequently, their levels are quantitative. As shown in Table 11, attributes can also be of nominal (on/off variables) or ordinal scale. A nominal attribute is the one where no natural order exists between the levels (eg with signal vs. no signal for mobile telephony). Selection of attribute levels in this case requires consideration as to what levels are likely to result in changes in preference (eg reception quality for mobile telephones). In the case of ordinal level attributes (eg voice quality in mobile telephony being (a) poor, (b) fair, (c) good) there is a natural order between levels and assignment of intermediate (additional) levels between extremes requires in-depth study of the underlying choice process through, for example, focus groups or cognitive testing.

### Step 3: Experimental design

Having defined the number of alternatives and the number of attributes and their levels, the next step in the design of a choice experiment is to construct a number of scenarios to be evaluated by respondents. For this purpose, choice modelling techniques are based on the concept of experimental design or the design of experiments. In particular, “*a designed experiment is a way of manipulating attributes and their levels to permit rigorous testing of certain hypotheses of interest.*” (Louviere et al, 2000).

---

<sup>8</sup> When the attribute is a quantitative (continuous) attribute

The number of possible scenarios can increase exponentially with the number of alternatives, the number of attributes and their levels. For example, consider the case of choosing between two modes of transport: bus and metro. Some common attributes may include price of ticket, travel time, frequency of service, comfort and reliability. Assuming that all attributes have four levels – though this is not always the case – it would be possible to generate  $4 \times 4 \times 4 \times 4 \times 4 = 1,024$  scenarios after considering all possible combinations among attributes and their levels. This is also known as a *complete factorial design* (Louviere et al, 2000). The design in its current form is suitable for *conjoint* analysis where respondents rate or rank each of the scenarios (Hensher et al, 2005). In the case of choice analysis where two or more alternatives are present then full enumeration of all possible choice sets is equal to:

$L^{MA}$  for labelled experiments, and  
 $L^A$  for unlabelled experiments,

where L is the number of levels, M is the number of alternatives, and A is the number of attributes. For example, a labelled experiment with three alternatives, five attributes at three levels each yields 14,348,907 possible scenarios (or treatment combinations) (ie  $3^{3 \times 5} = 3^{15}$ ). In the vast majority of choice modelling studies, it is impractical to use a complete factorial design unless the experiment involves a very small number of attributes or levels or both (Louviere et al, 2000) (see the discussion in *Step 4*).

The solution to the large number of scenarios produced by a complete factorial design comes from the field of statistical experimental design which provides the means to obtain significantly smaller subsets of all possible scenarios in a statistically efficient way. Therefore, the objective of the experimental design is to create experiments with a minimum number of scenarios while being able to infer the influence of all attributes when respondents choose, rate or rank alternatives or scenarios (Louviere et al, 2000). This objective leads to smaller choice tasks and the need for fewer sample points to support the estimation of the choice models.

The most frequently used experimental design is the *fractional factorial design*<sup>9</sup> and involves a selection or subset of the full factorial design in which the statistical properties are maintained in the best way possible (Louviere et al, 2000). Subsets of the full factorial design are selected using sampling techniques that result in practical and manageable designs with specific statistical properties. For example, fractional factorial designs can be *orthogonal* meaning that each of the attributes in the experiment has zero correlation with the rest of the attributes in the experiment or *balanced* in which all levels of each attribute occur with equal frequency (Bateman et al, 2002, Street et al., 2008).

However, all fractional factorial designs come with some loss of statistical information such as the ability to capture higher order effects known as *interactions* between two or more attributes (see, Louviere et al, 2000 and Chapter 4, for a discussion). Most studies develop main effects fractional factorial designs in which all two-way or higher order interactions among attributes are assumed to be insignificant. Louviere (1988) and Louviere et al (2000) argue that omission of interactions effects does not necessarily

---

<sup>9</sup> SPSS is a commonly used software package with a relatively straightforward module for the construction of orthogonal fractional factorial designs

lead to biased results as main effects account for as much as 90% of the explained variance.

Fractional factorial designs are in most cases suitable for the design of experiments (Hensher et al, 2005). However, over recent years, a number of other design approaches have been developed for discrete choice experiments enabling the generation of scenarios in a more systematic and statistically robust way. Johnson et al (2006) provide a summary of the full array of approaches to develop discrete choice experimental designs<sup>10</sup>. One of these approaches aims at developing *optimal* or *statistically efficient designs*. The objective of an optimal design is to extract the maximum amount of information for the choice task. In other words, *optimal designs* generate design matrices that can provide better estimates of attribute-level coefficients or ratios of attribute level coefficients than fractional factorial designs (Huber and Swerina, 1996, Bateman et al., 2002).

The latter is particularly relevant for WTP studies where the analyst seeks the ratio of an attribute level to the cost coefficient. The most common measure to optimise a design matrix is called D-efficiency (see, Huber and Swerina, 1996, Bliemer and Rose, 2006, Kuhfeld, 2009) and seeks to simultaneously satisfy four design properties: level balance, orthogonality, minimal overlap<sup>11</sup> and utility balance (see also, Hensher et al, 2005)<sup>12</sup>. It is now widely recognised by analysts that high quality designs play an important role in the accuracy of parameter and, hence, WTP estimates. As with many DCE issues, this area of experimental design is under continuous development (Louviere and Lancsar, 2009).

#### **Step 4: Choice set generation**

In this step, attribute-level labels can be assigned to numerical coding (ie attribute levels) which were used in the experimental design procedure. In the example presented in Table 11, if the attribute-level of 'price' takes the value of one then this will be replaced with the value of £20, whereas the label for the attribute 'colour' with attribute-level value equal to 2 will be replaced with the word 'yellow'. Therefore, any mapping between attribute levels and attribute-level labels is possible.

Problems may arise in the case of unbalanced designs where some attribute levels may occur more frequently than other levels. In such cases, Hensher et al (2005) point out that the analyst may prefer that a given attribute-level label be presented more than other attribute level-labels. Under this situation, a random allocation of attribute level labels to attribute-level codes is not possible.

The main objective in this step is to group the scenarios into choice sets. Scenarios can be presented individually (ie in a rating task), in pairs (ie in a pairwise-comparison task)

---

<sup>10</sup> SAS is a widely used software package for generating optimal design. A more recent, but equally powerful package tailored particularly to the design choice experiments is Ngene.

<sup>11</sup> Overlap relates to the statistical properties when pairing scenarios. A design has minimal overlap when an attribute level does not repeat itself in a choice set

<sup>12</sup> A choice set is utility balanced when the utilities of alternatives within each choice set are approximately equal. This can be rather problematic to address as it requires a priori knowledge of respondents' preferences. The analyst in this case needs to have some prior knowledge of the utility weights (coefficients) of the attributes when designing the discrete choice experiment. The rationale for this principle is to ensure that respondents are actually trading. Prior estimates can be obtained from a small pilot.

or in groups (ie in a ranking or choice task) according to the choice task that respondents will be asked to perform. However, the number of scenarios resulting from the experimental design can still be quite large and, hence, impossible for respondents to review and provide responses to all of them. For example, the full factorial design in the example of *Step 3* gave a total of 14,348,907 ( $3^{3 \times 5}$ ) scenarios while an orthogonal design matrix of  $3^{3 \times 5}$  combinations would result in 64 scenarios; still an excessively large number of scenarios to be evaluated by a single respondent.

Therefore, a judgement must be made between data efficiency (ie obtaining more observations per respondent) and respondent fatigue and boredom (which may result in biased and counter-intuitive estimates) about how many choice/rating or ranking tasks a respondent will be asked to perform (Bateman et al, 2002). In practice, respondents usually receive about eight choice scenarios (Carson et al, 1994); however, the number of scenarios to each respondent can range from 1 to 32 (Adamowicz et al, 1998b).

To avoid possible problems and keep the number of scenarios within a manageable size for respondents, Bateman et al (2002) propose three strategies:

- Reduce the number of attributes and/or number of attributes on offer
- Group the attributes into subsets (eg with common themes) and develop a smaller design for each set, or
- Split initial (large) designs into ‘blocks’ and offer each respondent only one block. Therefore, the 64 scenarios of the example above can be divided into eight blocks, and each respondent is given a set of eight scenarios from the same block (Hensher et al, 2005).

In many cases, it may be necessary to include an additional and fixed alternative in the choice set often called ‘opt-out’, ‘neither’, ‘non-participating’ or ‘status-quo’. By selecting this alternative, the respondent is allowed not to choose any of the alternatives on offer. Omission of the ‘opt-out’ option in some cases, may force respondents to choose alternatives unimportant to them or cause non-participation (Kjaer, 2005).

This is particularly important in cases where it is uncertain that the goods or services offered in the choice set will be preferred by all participants in the sample. Examples include the introduction of expensive and risky new technologies or new alternatives (Adamowicz et al, 1998b). Stated Choice models provide a theoretical basis for the inclusion and modelling of an ‘opt-out’ option through Random Utility Theory and discrete choice models (McFadden, 1973). Louviere and Woodworth (1983) have developed the conceptual and theoretical foundation for designing and implementing SC experiments in which non-choice is treated as another option faced by consumers.

Finally, the design and constructed choice sets need to be carefully tested through simulation, through the following steps (Bateman et al, 2002):

- Select some plausible values of the utility parameters
- Compute the utility of each alternative on offer for each respondent in the simulation sample
- Add the random error term sampled from an appropriate distributions
- Compute the implied choice, rating or ranking for each respondent

- Save attribute levels and implied choices to a data file, and
- Estimate choice models and test how well the initial assumed utility parameters are recovered.

## **Step 5: Development of the survey instrument**

### *Introductory text and presentation of choice sets*

It is important that respondents are first introduced to the choice task prior to asking them to complete the DCE. Typically, it is the first section in a stated choice survey that offers information about its context and provides instructions on how to complete the choice tasks. In particular, the aim of the introduction is to define the objectives of the study, the reasons for the choice of respondent and, also, stress the importance of respondents' participation (Adamowicz et al, 1998b). Moreover, the introductory text should mention the length of the survey and assure the confidentiality of responses.

The choice task itself is preceded by a set of standardised instructions on how to respond to the scenarios and detailed definitions of the attributes and their levels (Kjaer, 2005). In some cases, one or more 'warm-up' choice tasks are included so that respondents can familiarise themselves with the choice task (Carson et al, 1994). Hensher et al (2005) discuss in detail design considerations of the introductory section in a stated choice survey.

At the beginning of the choice task, a description of the given choice scenario places the available options into the context in which respondents should provide their preferences. Figure 10 shows an example of pre-choice text in a study examining the various aspects of patients' choices for earlier treatment (Burge et al, 2005). Framing of the presentation is very important and in some cases is in the form of a *descriptive story* (Hensher et al, 2005). Bennett and Blamey (2001) suggest that respondents must be aware that the good or service is part of an array of substitute and complementary goods and also that respondents may be reminded about their budget constraints.

Finally, if the questionnaire is about public funds, it is important that respondents are aware of competing agencies for public funds (Kjaer, 2005). The questionnaire should be structured in such a way that there is expectation that the information provided by respondents would be part of the overall decision making process. Otherwise, there is a risk that respondents would view the survey task as entirely hypothetical. While there is evidence that framing of the choice task influences the outcome of the experiment, it remains unclear how serious this effect might be in estimating welfare measures (Adamowicz et al, 1998b).

**WHICH HOSPITAL WOULD YOU CHOOSE?**

**PLEASE READ CAREFULLY**

We would like to find out what is important to you when choosing a hospital and treatment.

On the following 2 pages are a series of **possible** choices for the treatment for your current condition. We would like you to place yourself in the position where you were offered these choices. These choices are **imaginary** and **not real**, but we would like you to think about how you would act if you were in these situations. There are no right or wrong answers to these choices; we are only interested in your views.

In each case please indicate which option you would choose. We are aware that both options may be different to those you would like to be offered but we would like to know which you would choose if these were your options.

If you would not choose either of these options you may choose “neither”, but this should only be done if neither option would meet your needs and you would decide to either look for alternative treatment outside of the NHS or opt not to have your operation.

Please assume that the only differences between the hospitals are those presented, anything not mentioned would be the same at either of the two hospitals.

**Figure 10: An example of pre-choice text (Burge et al, 2005)**

### 3.4 General Design Considerations

This section focuses on bias issues around contingent valuation and discrete choice experiments. There are also more general market research design considerations such as sampling, methodology, sample size and questionnaire design. These and other market research design considerations are discussed in depth in Appendix B.

#### Source of bias

As mentioned in Chapter 2, SP methods are subject to a number of biases mainly arising from the hypothetical nature of the responses sought by these methods. Bias is defined “*as the difference between the distributions of hypothetical bids (or choices) obtained from a survey and the distribution of bids (or choices) that would be obtained in an actual demand revealing market setting*” (Schulze et al, 1996).

Table 12 summarises the different sources of bias for both CV and DCE and provides solutions to control for these biases where this is possible.

**Table 12: Definition of biases in Stated Preference methods (sources: Pearce and Ozdemiroglu, (2002); Hess et al, (2008); Louviere (2000))**

Bias	CVM	DCE	Recommendation
Hypothetical	Participants fail to take questions seriously. Therefore, in the majority of cases, respondents overstate their WTP for an increase in the quality of a good or service	Participants may not end up buying, switching to, selecting the product or service they stated	Ex ante correction: cheap talk script Ex post correction: market testing using revealed preference data
Information	Respondents have little, if any, prior experience with the proposed change in the good or service They also may make associations with other aspects that are not included in the scenario	Respondents may not be aware of all alternatives and their attributes including price They also may make associations with other aspects that are not included in the scenario	Ex ante: Provide respondents with a detailed and accurate description of the proposed change, product or service
Protest/ lexicographic/ strategic bias	Respondents may state a zero, positive or negative WTP because they want to protest for some aspect of the scenario Respondents provide specific answers in order to influence outcome	Respondents do not consider all attributes simultaneously, rather apply a heuristic where they define the preferred alternative by sequentially examining the attributes in the order in which they place most importance (known as lexicographic behaviour)	Ex ante: Control questions can be used to identify why the respondents provide such answers Ex post: Under certain conditions the analyst may choose to remove these observations
Non-Trader	N/A	Respondents always choose the same alternative without considering the values of the attributes	Ex ante: Control questions can be used to identify why the respondents provide such answers Ex post: Under certain conditions the analyst may choose to exclude these observations
Non-response	Respondents refuse to provide their valuation	Respondents refuse to provide their choice	Ex post: ICVM, analysis typically assumes that those who did not respond would not pay at all (zero value) In DCE, exclude observations. If possible weight remaining observations and test the difference between weighted and non-weighted observations in the sample
Fatigue / Frustration	Participants have to provide a number of responses eg iterative bidding questions	Respondent has too many exercises to respond to or dislikes all of the alternatives	Ex ante: careful survey design
Starting point bias	WTP anchored on initial stated value	N/A	Ex ante: Use open ended or payment card techniques
Framing effects	WTP depends on how the question is framed (negative or positive tone)	N/A	Ex ante: Possible presentation of questions in a "neutral frame"
Yea-saying  Nay saying	Respondent tries to please the interviewer Respondent concerned to counter the interviewer (only for iterative bidding or dichotomous choice methods with an interviewer – ie not self-completion)	N/A  N/A	Calibrate responses by scaling down (although the literature provides little guidance on the appropriate approach or scale of adjustment)

With regard to the DCE, more specifically, testing the validity and identifying bias in the choice experiment involves a number of tests. These tests are related to (Kjaer, 2005, Hensher et al, 2005):

- **Rationality** – Testing for rationality requires an internal consistency check which can be conducted either by introducing extra choice sets or the same choice sets twice. The extra choice set usually includes a dominant alternative. ‘Incorrect’ responses may be a result of irrational respondents, lack of understanding of the choice task or a simple mistake on the part of the respondent. In cases where a dominant alternative cannot be detected the analyst may include the same choice twice in order to examine the consistency of preferences as it is assumed that respondents will provide uniform answers to both choice sets
- **Framing** – Tests may involve varying the framing of the questionnaire for different subgroups of the sample in order to test for a framing effect
- **Ordering** – Test may include varying the order of attributes, choice sets, etc. for different subgroups of the sample in order to test the hypothesis of an ordering effect
- **Level range** – Varying the level range for different subgroups of the sample in order to test for level effects, and
- **Overlap** – Involves varying the amount of level overlap in a choice set for different subgroups of the sample in order to test for the effect of overlap on response variability and design efficiency.

### Uncertainty questions and cheap talk

The main criticism of the stated preference methods is their hypothetical nature. Respondents may overstate their willingness-to-pay while ‘pretending’ to buy a good or agreeing with a proposed policy measure. This phenomenon is called *hypothetical bias* (Louviere et al, 2000).

To overcome issues around hypothetical bias, analysts have developed *ex post* correction methods that involve calibration of WTP and the inclusion of a request that respondents consider their budgetary constraints. Calibration has been widely applied in CVM studies and involves a question which asks how sure the respondents are about their response to the WTP question (Kjaer, 2005). The response to the certainty question is usually recorded on a 10-point scale (Champ and Welsh, 2006).

The limitation of the calibration method lies in the transferability of the calibration function from one good to a different one. Fox et al (1999) note that “*calibration appears to be commodity specific and, as such, [the calibration approach] must proceed on a case-by-case basis until enough evidence exists to reveal any systematic bias that can eventually lead to a general calibration function*”.

The second approach focuses on adding to the questionnaire script a request that the respondent considers budgetary substitutes in responding to the valuation question. However, Loomis et al (1994) reported a number of problems encountered with this

particular approach and concluded that the approach of reminding respondents about their budgetary constraints is ineffective in removing hypothetical bias.

A more recent approach to correct for hypothetical bias is the 'cheap talk' approach. The advantage of the 'cheap talk' approach is that it provides an *ex ante* bias correction that can be applied to any evaluation (Lusk, 2003). The idea behind the 'cheap talk' approach is that a *cheap talk script* makes respondents aware of the problem of hypothetical bias prior to the actual stated preference task (Cummings and Taylor, 1999). In particular, respondents are faced with a discussion of what hypothetical bias is and why this may occur. Previous studies have found that inclusion of a 'cheap talk' script significantly reduces WTP making the valuation responses to hypothetical experiments indistinguishable from response to valuation questions involving actual payments (Cummings and Taylor, 1999). List and Gallet (2001) found that a 'cheap talk' script has the property of reducing WTP to unknowledgeable respondents while keeping knowledgeable respondents' WTP unaffected.

However, many questions remain unexplored regarding the application of 'cheap talk' scripts such as the interaction between knowledge and experience of respondents, payment vehicle and cheap talk effectiveness. Also, Carlsson et al (2004) found that the 'cheap talk' script significantly reduced the WTP values.

The DCE approach is less prone to hypothetical biases in WTP estimates than CVM as it replicates actual consumer purchasing condition more closely than CVM (Kjaer, 2005). However, testing hypothetical biases and the influence of cheap talk scripts in WTP estimates of DCE has been limited to only a few studies. The issue of hypothetical bias and its mitigation through 'cheap talk' scripts remains an open area of research and future work will determine whether such an approach is a valuable instrument in DCEs.

## **Perfect information and price elasticity effects**

An implicit assumption in choice models is that individuals possess perfect information on the alternatives available to them when making their choice. Therefore, individuals act in a rational manner drawing on this information. In practice, we know that this is unlikely to be the case, and choice models are typically specified to include error terms which amongst other things can account for noise within the choices that people are observed to make. In the case of stated preference survey data it is often assumed that this issue is less important as respondents are provided with the information on the available alternatives and are then asked to make their choices contingent upon this information ie there is less of an issue over the ambiguity of what respondents are considering when making their choices.

There is, however, one area where the assumptions of perfect information can become important within surveys seeking to elicit WTP and price sensitivity or switching between alternatives, and this is where responses to hypothetical choices are used to assume changes in demand from the current situation. Generally significant caution is required in estimating demand from stated preference data; however, in some situations this is necessary when other data sources are not available. In such cases it is informative to consider an intermediate stage within the questioning process that can be used to test the impact of improved information on current decisions ie:

- **Step 1** identify respondent's current choice between available alternatives;

- **Step 2** inform the respondent of the current prices of possible alternatives and check to see whether their choice would change under this improved information;
- **Step 3** vary the prices and investigate the impact on respondent's future choice making.

Many studies omit this second step and move straight to testing the impact of changed prices, which can in principle lead to biased estimates of *price elasticity* as the response incorporates a component which is not 'price driven', but rather 'information driven'. Table 13 provides a practical example of the proposed methodology described by the above three steps.

**Table 13: Example of providing improved information**

<p><b>Step 1</b> Previously, you mentioned that you pay £4.00 to travel to work by metro. Also you indicated that your total travel time from your home to work is 8 minutes that is 5 minutes travel time and 3 minutes waiting time.</p> <p><b>Step 2</b> The following list shows you current one-way single fares of all the available public transport options from your home to work area.</p> <table border="1"> <thead> <tr> <th></th> <th>Bus</th> <th>Tram</th> <th>Metro</th> </tr> </thead> <tbody> <tr> <td><b>Fare</b></td> <td>£2.00</td> <td>£2.00</td> <td>£4.00</td> </tr> <tr> <td><b>Travel time</b></td> <td>10 minutes</td> <td>10 minutes</td> <td>5 minutes</td> </tr> <tr> <td><b>Waiting time</b></td> <td>5 minutes</td> <td>4 minutes</td> <td>3 minutes</td> </tr> </tbody> </table> <p>Now could you please, tell me whether you would still travel by metro or would you choose bus or tram? - I would choose:</p> <p><input type="checkbox"/> Bus <input type="checkbox"/> Tram <input type="checkbox"/> Metro</p> <p><b>Step 3<sup>13</sup></b> Now I would like you to focus on the following situation in which travel times and cost have changed and tell me whether you would travel by Bus, Tram or Metro.</p> <table border="1"> <thead> <tr> <th></th> <th>Bus</th> <th>Tram</th> <th>Metro</th> </tr> </thead> <tbody> <tr> <td><b>Fare</b></td> <td>£2.50</td> <td>£3.00</td> <td>£4.50</td> </tr> <tr> <td><b>Travel time</b></td> <td>8 minutes</td> <td>7 minutes</td> <td>5 minutes</td> </tr> <tr> <td><b>Waiting time</b></td> <td>5 minutes</td> <td>3 minutes</td> <td>2 minutes</td> </tr> </tbody> </table> <p>- I would choose:</p> <p><input type="checkbox"/> Bus <input type="checkbox"/> Tram <input type="checkbox"/> Metro</p>					Bus	Tram	Metro	<b>Fare</b>	£2.00	£2.00	£4.00	<b>Travel time</b>	10 minutes	10 minutes	5 minutes	<b>Waiting time</b>	5 minutes	4 minutes	3 minutes		Bus	Tram	Metro	<b>Fare</b>	£2.50	£3.00	£4.50	<b>Travel time</b>	8 minutes	7 minutes	5 minutes	<b>Waiting time</b>	5 minutes	3 minutes	2 minutes
	Bus	Tram	Metro																																
<b>Fare</b>	£2.00	£2.00	£4.00																																
<b>Travel time</b>	10 minutes	10 minutes	5 minutes																																
<b>Waiting time</b>	5 minutes	4 minutes	3 minutes																																
	Bus	Tram	Metro																																
<b>Fare</b>	£2.50	£3.00	£4.50																																
<b>Travel time</b>	8 minutes	7 minutes	5 minutes																																
<b>Waiting time</b>	5 minutes	3 minutes	2 minutes																																

<sup>13</sup> Scenario developed through a discrete choice experiment

## Follow-up questions

Following the completion of the choice task, it is possible to include follow-up questions in order to try to understand the reasoning behind respondents' choices (or preferences). These questions can relate to other validity aspects such as misunderstanding on the part of the respondent and response aberrations (Bennett and Blamey, 2001):

- Ability to understand and the degree of difficulty of the choice task and the questionnaire in general
- The examination of reasons for the observation of dominant preferences (lexicographic preferences, decision-making heuristics)
- The examination of the cost attribute: existence of protest bias, payment vehicle bias, strategic bias
- Uncertainty question (ie “*Are you sure about your decision regarding how much you would pay?*”).

### 3.5 Enhancing SP Information: Combining Stated with Revealed Preference Data

As mentioned in **Chapter 2**, revealed and stated preference approaches have both advantages and weaknesses. Historically researchers have seen revealed and stated preferences as substitutes when considering the choice of valuation methods. Empirical evidence about the validity of stated preference methods (Cummings et al, 1986), and the fact that the strengths of revealed preferences are the weaknesses of stated preference approaches have led to a new paradigm for combining and jointly estimating revealed and stated preference data (Bradley and Daly, 1997, Whitehead et al, 2008).

While combining revealed with stated preference data could be challenging for CC's merger studies due to their tight time-schedules, this approach can help in addressing the weaknesses of both RP and SP data, in particular by:

- **Exploring new markets, products and services.** RP data are limited to analysing behavioural response within an existing market and attribute range. For example, the use of RP data can be of limited use when policy interventions, services or products are new. SP surveys can be designed to collect data on this hypothetical behaviour and SP data allows an examination of behaviour beyond the range of attributes in the current market. Combining RP and SP data allows an extension of the behavioural model beyond the limited range of existing attributes or goods (Ben-Akiva et al, 1994).
- **Taking advantage of the best features of the two approaches.** RP data offer real market observations, but lack information about new products or services. On the other hand, SP data offer a solution to the latter issue, but remain hypothetical and do not take into account budget and other constraints on behaviour. SP surveys can be used to understand changes in participation and market size with new products or service quality change (Hensher and Bradley, 1993, Whitehead et al, 2008). Combining RP and SP data is also a way of grounding and enhancing SP

estimates in actual behaviour, while extending the range of goods or services of interest beyond what is currently observed.

- **Improving the statistical efficiency of attribute-valuation estimates.** Even when there are existing data (RP) about the product or service under consideration, there is often high correlation between the variables of interest and other variables (eg vehicle efficiency and vehicle type) or there may exist endogeneity between attributes and respondent or other characteristics (eg choice of vehicle route to reduce the risk of accident and driver's experience) (Cameron et al, 1996). Combining RP and SP data under an appropriate experimental design can help control for correlation and avoid endogeneity problems (Von Haefen and Phaneuf, 2008).
- **Providing a means of more efficient and cost-effective data collection.** RP cross-sectional studies collect only one data point per respondent. Also, panel surveys or cross-section time series surveys collect multiple observation points; however, individual preferences involve different conditions over time, and therefore allow for time changes in behaviour. Moreover, panel surveys may encounter other issues such as loss of respondents, which lead to sample selection problems and increased costs. SP questions, on the other hand, can help increase sample size and statistical efficiency in model estimation. As a result, by enhancing RP with SP data, the analyst can achieve model efficiencies with a smaller sample size, thus reducing the overall costs of a study (Whitehead et al, 2008).
- **Offering a check on convergent validity and reliability of both RP and SP methods.** Hypothetical bias is a key problem with SP data. Combination of SP with RP data can help to mitigate this problem as well as assess the validity and reliability of SP data. Carson et al (1996) propose that comparisons between contingent valuation and revealed preference estimates are generally assumed to represent tests of converged validity.<sup>14</sup> These tests are possible when two or more measurement techniques are potentially capable of measuring the desired quantity, but they do so with error. Converged validity exists when two different methods yield measures (eg WTP) that are not statistically different. Moreover, RP methods such as the Travel Cost have been criticised for the measurement error and endogeneity of trip cost (Randal, 1994). Combination with SP data can help in isolating the effect of changes in trip costs (Englin and Cameron, 1996). Therefore, joint estimation of RP and SP data can be used to validate both types of data.

The modelling approaches for combining RP and SP information are classified based on the type of the dependent variable under study (ie continuous or discrete), and whether stated choices or contingent valuation has been used to obtain the SP data. Table 14 provides a set of criteria for selecting across the available approaches. In general, joint estimation of RP and SP random utility models are more appropriate when the main policy impact is on substitution between alternatives or changes reflected into more than one attribute. This approach can be quite complicated. Examples and discussion using

---

<sup>14</sup> Cummings, Brookshire and Schulze (1986) emphasised that revealed preference data should not be treated as the "truth" measure, but as random variables, which are sensitive to the attributes of the good or service, the functional form used in estimation, and other technique-specific assumptions. (Cummings, R.G., Schulze, W.D., Gerking, S. & Brookshire, D.S. (1986) Measuring the elasticity of substitution of wages for municipal infrastructure: A comparison of the survey and wage hedonic approaches. *Journal of Environmental Economics and Management*, 13, 269-276.)

this approach can be found in Hensher and Bradley (1993), Bradley and Daly (1997), and Cherchi and Ortuzar (2002).

Continuous choice models use RP data on variables such as number of trips or number of miles driven. These data are combined with SP data in which respondents make similar choices under hypothetical conditions. As in discrete choice models, the data have the same structure and can be pooled to produce joint model estimates (Whitehead et al, 2008). This approach has found wide applicability in recreation demand models where respondents provide information on how many trips they make to the recreational site (or group of sites), and the cost to them of visiting the site. In this context, hypothetical questions focus on how the trip making behaviour of each respondent would change if costs of visiting the site changed.

**Table 14 Potential combinations of revealed and stated preference data (sources: Bateman et al (2002); Whitehead et al (2008))**

Focus of policy impact	RP Data	SP Data	Modelling Approach
Substitution between alternative goods or services/ Changes in multiple attributes	Observed choices and characteristics of chosen goods or services including price	Stated choices through discrete choice experiments	Random utility approach - discrete choice modelling (MNL, Nested MNL, Mixed MNL)
Total demand for a good or service	Data on continuous variables (eg number of trips to recreation sites, number of vehicle-miles driven)	Contingent behaviour data similar to RP data, but in hypothetical conditions	Continuous choice models (eg Regression, Poisson, Tobit, Negative binomial)

While the combination of RP and SP data presents a number of advantages over the use of RP or SP data alone, a number of challenges include (Bateman et al, 2002):

- Model complexity and availability of user-friendly software to implement the combination
- How to reconcile potential differences between stated behaviour and actual behaviour in developing predictive models for new scenarios
- Context-specific applicability, as there might exist an RP technique that provides the solution to the problem
- Empirical evidence in the estimation of combined RP and SP models is limited. Whitehead (2008) discuss a number of challenges with regard to combined RP/SP models
- Longer and intensive questionnaires may result in lower response rates, lower quality of responses, etc.

## 4. APPRAISAL OF SP VALUATION TECHNIQUES IN PRACTICAL APPLICATIONS

Chapter 2 briefly discussed the theoretical differences between contingent valuation and the discrete choice experiment approaches. Having now seen the different design steps involved for design valuation questions or scenarios using both methods, this section discusses the issues around the practical implementation of these methods.

In practical applications, the analyst has to balance between the rigour of the method and the time and budget available to conduct a valuation study. As shown in Table 15, discrete choice experiments take much longer to implement as they require background study, focus groups and testing of the potential attributes and levels prior to the official commencement of the survey. Following pilot testing, it might be necessary to revise the experimental design should any of the attributes or their levels require amendment. The experimental design stage is the most important task in the design of DCE and poses more challenges than the procedure of designing valuation questions in CVM. At the conceptual stage, therefore, the time and budget available are the two most important factors in deciding whether a CVM or a DCE would be preferable.

Time and resource availability issues should be taken into account along with the objectives of the valuation question, and specifically whether it is the total value of the good or particular attributes of the good being under investigation. Typically, the CVM approach provides a description of changes in goods or services, while assuming that information about all alternative and substitute goods is available. However, the analyst has no direct evidence of that assumption being true. Respondents are required to answer a question which involves paying for the improved good or service (Boxall et al, 1996). A problem with this approach is its reliance on the accuracy of the information, and that any errors in the information discovered after the fact cannot be changed. On the other hand, the DCE approach relies on the representation of a choice situation using a set of attributes rather than describing a specific change in the good or service.

The ability in DCE to present the good or service in question within a 'package' of potential trade-offs is a way of minimising information transfer and other potential biases, and modelling realistic choices (Kjaer, 2005). This is very important when the good in question is not familiar to respondents. Therefore, the DCE approach relies more on the accuracy and completeness of the characteristics used to describe the good or service than any particular description of the good or service as a whole (Boxall et al, 1996). Moreover, cost in DCE is just one of the several attributes presented to describe the good or service in question. Thus, the emphasis on cost is less and CVM related biases may be reduced. Overall, the DCE is the recommended approach in cases which valuation of one or more attributes of a good or service are of particular interest.

**Table 15: Practical assessment between CVM and DCE**

Stage		Contingent valuation method	SP Discrete choice experiment
<b>Conceptual</b>	Cost	Low-medium	High
	Timing of the valuation exercise	Short (1-2 months)	Medium-long (3 - 6 months)
<b>Design of the valuation task</b>	Valuation of the total good or individual attributes	Total Good	Individual attributes
	Complexity for designing the valuation scenario	Low - medium	High
	Specialised software and analysts	Low	High
<b>Preference Elicitation</b>	Task complexity (time required in explaining the task to respondents)	Low-Medium	High
	Avoids yea-saying (compliance bias)	No	Yes
	Stability of preferences	High	Medium
<b>Analysis and Results</b>	Modelling expertise and software requirements	Low	Medium-high
	Estimation of marginal effects and attribute values simultaneously	No	Yes

With regard to the preference elicitation task, and as discussed in Section 3.3, selected attributes and levels are used along experimental design procedures to construct alternatives that reflect different versions of the good or service under investigation. Individuals are asked to choose from a ‘choice set’ constructed from different packages (Hensher et al, 2005). Experimental designs can be constructed using a range of different approaches including traditional orthogonal-design matrices and optimal/statistically-efficient designs. The former can be developed using widely available statistical software such as SPSS. Efficient designs, on other hand, require advanced experimental-design software such as SAS<sup>15</sup> and Ngene<sup>16</sup>, which may add a significant amount to the budget in terms of time and resources to develop the design and, subsequently, monetary costs.

Potential savings in the use of efficient designs would come from the smaller sample required to estimate robust attribute-weights (Bliemer and Rose, 2005). Experimental design is not part of the overall design procedure in CVM questions. While this component makes the DCE design procedure significantly different and more difficult than CVM, it has important implications in the quality and robustness of the valuation study.

Table 15 also shows that there are both advantages and drawbacks in the preference elicitation tasks between DCE and CVM. A DCE, for example, is more demanding in

<sup>15</sup> <http://support.sas.com/techsup/technote/ts723.html>

<sup>16</sup> <http://choice-metrics.com/>

terms of cognitive burden as respondents may be asked to make a choice among two or more alternatives each described with, typically, 5-10 attributes (Carlsson and Martinsson, 2003). Moreover, preferences may be unstable throughout the experiment giving rise to lexicographic and non-trading behaviour, which can have significant impact on model results (Hess et al, 2008).

Carlsson and Martinson (2003) argue that instability in choices may be due to learning effects and suggest that a set of ‘warm-up’ choice questions should be given to individuals prior to the actual choice tasks. On the other hand, the advantage of choice experiments is that it avoids yea-saying behaviour since respondents are not faced with the stark ‘all or nothing’ choice in the design of CVM (Ready et al, 1996, Hanley et al, 1998). Finally, the multi-attribute evaluation information that is measured in DCE could be elicited using repeated CVM questions. However, a large number of CVM type questions would be needed, and it would be difficult to maintain some degree of orthogonality in the design and administration of such an experiment (Adamowicz et al, 1998a).

With regard to the analysis of the collected data, CVM and DCE also exhibit different levels of complexity, which corresponds to different needs for analytical skills, software and computing resources. The analytical methods for CVM include regression, logistic-regression and panel data models, which can be developed using standard statistical software (eg SPSS). Attribute weights in DCE can be estimated using different *discrete choice models* based on Random Utility Theory (eg multinomial logit, nested logit, mixed logit, GEV family models) that vary in complexity and computational requirements (McFadden, 1973, Ben-Akiva and Lerman, 1985, Hensher and Greene, 2003) and require specialised econometric software (eg Alogit<sup>17</sup>, Biogeme<sup>18</sup>, and LIMDEP/NLogit<sup>19</sup>).

The information from DCE can be particularly useful for those involved in policy decisions and setting resource allocation priorities since they can be used to produce market shares and allow testing of policies affecting demand (Louviere et al, 2000). Moreover, the added value of DCE over CVM is that estimated weights provide information on the relative importance of a number of selected attributes in the experiment. Practically, it is standard in DCE to assume a linear utility function in which “*the value of the whole is equal the sum of different parts*”.

However, this raises two concerns; firstly, regarding the influence of attributes not included in the experiment (and thus not included in the utility), and whether the linear additive assumption is appropriate (Kjaer, 2005). Foster and Mourato (2003) conducted a systematic comparison of dichotomous choice DCE and CVM with emphasis on embedding/scope effect and the linear additive utility function assumption. Their study examined the social value of services provided by the charitable sector in the UK – such as that related to housing and homelessness, social services and health research. The results showed that the value of the good as a whole was larger in the DCE than in the CVM, and thus suggested that respondents perceive CVM and DCE differently. Furthermore, Foster and Mourato (2003) suggested restricting the use of DCE to

---

<sup>17</sup> <http://www.alogit.com/>

<sup>18</sup> <http://transp-or.epfl.ch/page63023.html>

<sup>19</sup> <http://www.limdep.com/>

evaluation of a single isolated change of policy that is part of a larger set of changes or policies.

In practice, the choice of the method would first of all depend on the context of the experiment and second on the trade-off between the advantages and disadvantages of the two methods (Carlsson and Martinsson, 2003).

## 5. REVIEW OF THE CURRENT PRACTICE

### 5.1 Introduction

This chapter reviews five recent Competition Commission studies in order to explore how the current practice of CC compares with the current practice reported in the academic literature and other competition commissions worldwide. The following case studies were suggested by the Competition Commission as they represented a range of approaches.

- Eastman Kodak Company & Colour Care Limited (2001)
- First Group Plc / ScotRail Railways Limited (2003)
- National Express Group PLC/ Greater Anglia Franchise (2004)
- Acquisition of the Greater Western Passenger Rail Franchise (2005)
- Stagecoach / Scottish Citylink (2006).

For these studies the Competition Commission provided a range of files for each which variously included the brief, the proposal, questionnaires, reports, data, statements from interested parties and the Competition Commission.

The next section of this chapter reviews the valuation questions used in these studies. Section 5.3 presents an illustrated example of moving from a CVM to a DCE study.

In Section 5.4 we review current practice in three other competition authorities, based on interviews with representatives of the US Department of Justice, the EU Directorate General for Competition and the US Federal Trade Commission.

Finally, Section 5.4 sets out our recommendations for future Competition Commission surveys.

### 5.2 Review of Pricing Questions Used by the Competition Commission

#### **Types of valuation questions in studies by the Competition Commission**

For the purposes of this report, the study team has also reviewed the hypothetical questions from five studies carried out as part of merger inquiries, which were provided by the Competition Commission. Mergers involve short studies with duration of a total of 12 weeks, including 1.5 weeks procurement, 4-5 days consultancy, 6.5 weeks for market research with survey results obtained at the beginning of week 10.

This review has further classified these studies based on the type of hypothetical questions. In particular, we have identified two broad categories of hypothetical questions: *open-form* and *closed-form*.

*Open-form* questions ask respondents to state the price at which they would stop using their current mode of travel or switch across different modes of public transport. This group of open-form studies has been further classified based on whether switching concerns own mode, between modes or across all possible public transport modes.

In *closed-form* questions, respondents are offered products or services at different prices or service-level/price bundles (eg film processing time and price) and are asked to make a choice. In addition, this category includes discrete choice experiments, in which respondents are offered a number of alternatives (packages of attributes), which are constructed based on an experimental design. Similar to the other close-ended questions, respondents are asked to choose one of the available alternative options. Different types of closed form questions identified across studies are summarised in Box 6.

#### Box 5: Open-form questions

- 1. Open ended price/time (own mode)**  
What is the maximum journey time/cost that you would accept before you stop using this mode of transport?  
(First Group plc / ScotRail Railways Limited 2003)
- 2. Open ended price/time (between modes)**  
For a similar ticket on the bus, what price of ticket would make you use the bus instead of the train for that journey?  
(First Group plc / ScotRail Railways Limited 2003)  
(National Express Group PLC / Greater Anglia Franchise 2004)
- 3. Open ended price/time (all PT modes)**  
Is there a fare for this journey by any mode or transport that would have been just enough to stop you using public transport on this route?  
(National Express Group PLC / Greater Anglia Franchise 2004)

#### Box 6: Closed-form questions

- 1. Menu of priced items and choice of one**  
If these were the prices which of the following services would you normally choose?  
(Merger: Eastman Kodak Company and ColourCare Limited 2001)  
Would you have used *your next best alternative* instead of the coach service you travelled on if the price of your coach ticket had been ORIGINAL PRICE +50% instead of ORIGINAL PRICE?  
(Stagecoach / Scottish Citylink 2006)
- 2. Menu of responses to price change and choice of one**  
If all fares went up by 5% (that is, 5p in every £), what would you be most likely to do?  
(Greater Western Passenger Rail Franchise 2005)
- 3. Discrete choice experiments (multiple attributes)**  
We would now like you to imagine that the coach and train journeys available to you were as described in the scenarios below. In each case, please consider what decision you would make if you were faced with the times and costs presented in each scenario  
(National Express Group PLC / Greater Anglia Franchise 2004)

### Review of hypothetical questions in CC studies

From a methodological point of view, the hypothetical questions in CC's studies are either in the form of contingent valuation questions or discrete choice experiments. Contingent valuation questions usually are developed using a mix of methods (eg dichotomous choice and iterative bidding) whereas discrete choice experiments are developed using standard SP methodology.

Table 16 summarises the key characteristics of each of the studies including objectives, the attributes of particular interest, and the types of hypothetical questions that were

asked. All studies focus on hypothetical cases involving potential change in competition at the local or national level, and therefore of particular interest is changes in the behaviour of customers with regard to changes in the supply of services or products. Prices and quality characteristics of products or services are of particular interest.

**Table 16: Summary of CC merger case studies**

<b>#. Study (Year)</b>	<b>Objective(s)</b>	<b>Attributes</b>	<b>Type of question</b>
<b>1. Eastman Kodak Company &amp; Colour Care Limited (2001)</b>	Explore likely changes in behaviour should there be an increase in the price of the developing and printing service. Examine, in particular, whether consumer would switch between services for 10% increase in the retail price of next-day-to-within-a week developing and printing service	Price - Location where consumers would like films to be developed & printed (not included in the bidding) - Speed	Mixed mode: Menu-based choice of price-service bundles (conjoint type of design) - Prices increase in three out of five attributes
<b>2. First Group Plc / ScotRail Railways Limited (2003)</b>	Factors influencing people to use bus (more or less), factors that make people switch from bus to train and vice versa, price sensitivity	Choice between Bus and Train - Journey time - Acceptable frequency - Ticket cost - Available alternatives per respondent	Price, frequency of service, continued use of the service. Open-form questions about switching between two modes of public transport (ie bus, train). Identifying thresholds of switching
<b>3. National Express Group PLC/ Greater Anglia Franchise (2004)</b>	Examining the competition between London and Southend lines  Extent of competition between coach and train services in Greater Anglia	Switching between train and coach; pricing question about stopping using public transport - Time on vehicle - Service frequency - Time to and from stations - Cost (daily one way)	Open-form contingent valuation questions and Stated Choice Experiment
<b>4. Acquisition of the Greater Western Passenger Rail Franchise (2005)</b>	Examine the effects of the acquisition on certain routes in the area covered by the Greater Western rail franchise	- Opinion on the increase of the ticket price over the last 12 months - Opinion on whether the price of a ticket for other modes of transport connecting the points of travel decreased, increased or remained the same - Intended change in behaviour (mode choice) in case existing train fare increased by 5% (or 5p in every £)	Closed-form questions with only one being a pricing/intended behaviour questions

Table 25 in Appendix C sets out the research method used for each study.

This section highlights the potential advantages of using the hypothetical questions of each study and identifies points in the questions which need further consideration since they are likely to result in biased and misleading results.

## **1. Eastman Kodak Company & Colour Care Limited (2001)**

The objective of the study was to examine the effect that a potential merger would have on consumer film processing behaviour. The focus was on adults aged 15 years or older in Great Britain, and particularly those who used a next day to a week services for APS and regular colour negative film. The preference information came from questions 5 to 10 in the questionnaire:

- Q5. Here are some typical current prices for the various ways of processing a 25-exposure APS film. If these prices were to apply everywhere in the country, which service would you normally choose?
- Q6. If the prices of the next day, 3 day and 6 day services were to increase by 76 pence, but the prices of same day and mail order remained the same, which would you choose? These new prices are given on the right of the card.
- Q7. Now looking at this card, what if there was a further price increase of 74 pence for the next day, 3 day and 6 day services, but the prices of the same day and mail order STILL remained the same which would you NOW choose?
- Q8. Here are some typical current prices for the various ways of processing a regular 24 exposure colour film. If these prices were to apply everywhere in the country, which service would you normally choose?
- Q9. If the prices of the next day, 3 day and 6 day services were to increase by 50 pence, but the prices of same day and mail order remained the same, which would you choose? These new prices are given on the right of the card.
- Q10. Now looking at this card, what if there was a further price increase of 50 pence for the next day, 3 day and 6 day services, but the prices of the same day and mail order STILL remained the same which would you NOW choose?

Hypothetical menu-based questions were used to examine individuals' preferences for film development and processing option when the prices of next-day, 3-day and 6-day processing and developing times increased by 50%. The co-linearity in the prices in the choice task restricts the analysis and does not allow the analyst to recover differences in price response across the individual products. To avoid this limitation, an alternative could have been the development of a DCE using a fractional-factorial design matrix as described in Section 5.3.

Moreover, the survey asked for other factors which influenced the choice of different developing and processing times including location, trust for mail order delivery and which were not taken into account in the analysis of preferences. A univariate analysis of the proportions in the sample provides a limited evidence base upon which to draw conclusions such as younger people were more concerned with speed and were more likely to choose faster services. The ideal option to support this hypothesis would be to design an unlabelled DCE where processing and developing times, price, location and

other characteristics would vary in the experiment. An example of the proposed unlabelled DCE with three attributes is given in Table 17 and Figure 11.

**Table 17: Attributes and levels in an unlabelled DCE**

Attribute	Levels
Speed of return	Same-day, next-day, 3-day, 6-day, Mail-order
Price	£4.75, £5.49, £6.49, £7.25, £7.49, £8.25, £8.99
Location	At your local store, only in the City Centre,

	Option A	Option B	Option C
Processing speed	Same day	Next-day	6 days
Price	£8.99	£6.49	£4.75
Location	At you local convenience store	Only in the city centre	At you local convenience store
I would choose	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Figure 11: Example of an unlabelled choice exercise involving different factors that may influence choice**

Using the above mentioned DCE (see Table 17 and Figure 11) it would be possible to examine the effect of price, processing speed and location on choice independently from each other. In addition, individuals' WTP could be computed, for example, for processing speed and location of film processing and development. Finally, further analysis would allow analysis of preferences across different socio-economic groups eg by age. Different constraints during the experimental design could be used to prevent meaningless scenarios eg mail-order being more expensive than same-day.

## 2. First Group / ScotRail Railways Limited 2003

This study aimed to examine the attitudes and behaviours of First Group's customers on the sections of bus routes between stopping points that were also served by ScotRail trains (ie overlapping services) around Glasgow and central Scotland. Survey participants were those who had made a recent journey from home on local public transport where they had a choice of using either bus or train for the main part of their journey.

The hypothetical questions used in this study were in sections 4 and 5 of the questionnaire. Section 4 implicitly assumed that respondents possessed full information of all different options. Specifically, respondents were asked to report price, frequency of service, travel and walking to station/stop time, reliability, total travel time, etc of the mode they would use instead of their current mode of travel. However, there may have been respondents who were not aware of all alternatives available and their characteristics (eg travel time, frequency, etc).

An alternative approach would be for the interviewer to provide the relevant information to respondents and asked them to state whether they would choose the bus instead of train and vice-versa (see Section 3.4). In such cases, a judgement is required as to whether it is desirable to understand respondents' behaviour under their current information, or their behaviour in the presence of more accurate information. This requires a judgement as to whether individuals would be likely to seek better information in the presence of a service change to inform them of the alternatives that they could consider. A third alternative would be to ask respondents whether they

would seek out new information and to present this only to those who state they would do so.

Section 5 is entitled hypothetical switching; however, the focus was on respondents' tolerance regarding different travel characteristics including: maximum journey time, minimum service frequency and maximum price of the ticket respondents would pay before they would stop using their current mode of public transport (eg switching from bus to train or from train to bus). In all cases, it was assumed that only one travel characteristic (eg price) varied (respondent chooses the minimum price) while other characteristics (eg frequency, travel time) remained the same. As was also suggested by CC's final report, the variability of these attributes was limited and so no conclusions could be drawn with regard to (own and cross) elasticities of different modes.

A better option may have been to have included an additional section with a DCE exploring trade-offs for the attributes of interest for both bus and train (see Figure 12). Once the model has been developed different pricing, level of service, type of ticket scenarios could be used to identify market shares of the two modes.

	Bus	Train
<b>Walking time from home to bus stop or station</b>	10 min	15 min
<b>Frequency of service</b>	6-10 minutes	1-5 minutes
<b>Travel time</b>	30 min	20 min
<b>Reliability</b>	Quite reliable	Very reliable
<b>Time to reach the final destination</b>	10 min	15 min
<b>Number of interchanges</b>	2	0
<b>Waiting time for connections</b>	2 min	0 min
<b>Type of ticket</b>	Single journey	Return journey
<b>Price</b>	£2.00	£6.00

Figure 12: Examples of DCE in the context of First Group/Scott Rail (2003) study

### 3. National Express Group PLC Greater Anglia Franchise 2004

This study was part of the inquiry into the acquisition by National Express Group of the Greater Anglia rail franchise, and was based on two consumer surveys:

- Train passengers travelling on the C2C and One Great Eastern lines between London and stations in the Greater Southern area,
- Train passengers travelling on One Anglia services between London and Ipswich, Colchester and Norwich and amongst National Express coach passengers on similar routes.

The objectives of the study were to examine:

- the extent to which the two train lines between London and Southern compete with one another, and
- the extent to which coach and train services in the Greater Anglia area compete with one another.

**c2c - One Great Eastern lines** – The survey asked respondents if they were aware how c2c and One Great Eastern services compare in terms of journey time and price. Respondents who were aware of the alternative services provided further travel

information including time and mode to reach the station, travel time and destination. On the other hand, respondents who were not aware of the alternative route were not asked further questions and were directly given the stated choice experiment. While the study correctly identifies respondents who had some or full information about the characteristics of the alternative journey, it provided no additional information to those who were not aware of the alternative route. Therefore, there might be information bias in the sample, which does not seem to be controlled in the choice models.

The choice experiment followed a standard methodology and allowed respondents to opt-out from the choice between c2c and One Great Eastern in case none of the options was suitable for the respondents. At the end of the stated choice experiments it would have been useful for the analysis to include a number of follow-up/diagnostic questions that would further identify individuals who had difficulties in understanding the choice experiments and might lead to hypothetical bias.

**One Anglia and National Express coach** – In contrast to the c2c-One Great Eastern Lines questionnaire, the survey assumed that respondents were aware of the alternative route. Therefore, there is again the issue of information bias as respondents might have never used or known the characteristics (cost, stations, travel times, etc) of the alternative route by coach.

The stated choice experiment followed a standard methodology. There were, however, ‘easy’ options included in the experiment. Figure 13 shows an example where the choice of the ‘train’ alternative was expected to be chosen. Specifically, the prices of coach and train were very similar, while the ‘coach’ service frequency and duration were much worse than the ‘train’ option. In practice, ‘easy’ options should be further examined and possibly be excluded from the experiment since they offer little additional information to the analyst and are likely to drive respondents into providing meaningless responses. Alternatively, a decision may be taken to include them as "warm-up" questions or internal consistency checks, although this does bear the risk that respondents discount the exercise as poorly designed. Finally, a number of diagnostic questions would have helped to identify whether respondents understood the choice task and provided reliable responses.

<b>SCENARIO</b>	<b>Coach</b>	<b>Train</b>
Time on vehicle.....	3 hours 10 minutes	2 hours
Service frequency.....	Every 3 hours	Every 30 minutes
Time to and from stations.....	As now	As now
Cost (Daily one-way).....	£12.00	£13.00
<b>In this situation I would choose (tick box):</b>		
<b>To use neither</b> <input type="checkbox"/> 1	<b>Travel by coach</b> <input type="checkbox"/> 2	<b>Travel by train</b> <input type="checkbox"/> 3

**Figure 13: Example of an ‘easy’ option in the DCE**

#### **4. Greater Western Passenger Rail Franchise 2005**

The Greater Western Passenger Rail Franchise 2005 study aimed to establish if the acquisition of the Greater Western Passenger Rail Franchise by First Group plc, National Express Group plc and Stagecoach Holdings plc, would result in a substantial

lessening of competition in the market for the supply of passenger transport services in the franchise area. The specific research objectives of this study were to:

- Understand attitudes and the propensity to use alternative transport links available in the franchise area
- Identify reasons for choice of transport services
- Identify factors for current/recent journey
- Establish whether passengers consider possible services as realistic alternatives.

In contrast to the studies mentioned above, the questionnaire collected information on travellers' awareness about the availability of train/bus service connecting the points of travel on the day of the interview (Q13). However, with a computer-based or interviewer-led survey there could have been an additional section describing the different alternatives and their characteristics eg price, frequency, etc to respondents who were not aware of other available options. At that stage, respondents with little or no information were treated the same with those who were aware of the characteristics of the alternative options. Therefore, responses obtained from Q18 (*"If First Group/Wessex Trains no longer operated a bus/train service on this route, what would you be most likely to do instead?"*) may be influenced by information levels. An example on how respondents could obtain more information during the survey has already been discussed in section 3.4.

The hypothetical question Q24 (*"Please think about the fare you have paid to First Group/Wessex Trains for your bus/train ticket today, (including season tickets if relevant). If all of First Group's bus/Wessex Trains fares went up by 5% (that is, 5p in every £), what would you be most likely to do - or would you not permanently change your travel behaviour in any of the ways below?"*) could be replaced with a discrete choice experiment which would involve the most popular choices: (a) First Group/Wessex Trains, (b) Another bus company, (c) car (passenger or driver), (d) Coach, etc. The layout of the choice task could be similar to those suggested in case studies 1 and 2 above. Hypothetical scenarios could involve different situations of price changes ( $\pm 5\%$ ), time delays, travel times etc. Following the development of the discrete choice model, hypothetical scenarios would enable the researcher to examine the market shares of different modes should the First Group/Wessex Trains alternative have been available.

## **5. Stagecoach / Scottish Citylink 2006**

The objectives of this study were to:

- Establish bases for estimates of the composition of passenger flows by journey purpose; means of reaching pick-up point; key demographics; and ticket details;
- Identify the factors in choosing mode of travel and supplier that are most important to passengers;
- Establish the extent to which passengers are aware of other travel options;
- Rank passengers' other travel options in order of attractiveness;
- Test passengers' stated intentions in response to changes in ticket prices and service characteristics;
- Gather demographic information for the analysis of survey responses by different passenger groups.

The hypothetical questions focused on the propensity of respondents to change their travel arrangements in responses to (hypothetical) changes in prices and levels of service. For example, Q13 in the follow-up CATI survey (“*Firstly, would you have [ANSWER AT Q11: eg. use a bus] instead of the coach service you travelled on if the price of your coach ticket had been [ORIGINAL PRICE +50%] instead of [ORIGINAL PRICE]?*” IF ‘NO’, GO TO Q15. IF ‘YES’ ASK FOR +1% INCREASE) offered respondents different price increases (a combination of iterative bidding questions and dichotomous choice) ranging between 1% and 50% and asked them whether they would change their preference if the Citylink/Megabus alternatives had not been available. As mentioned in the commentary on the case studies above, the results of these questions may be influenced by how much information respondents had on the services and their characteristics (the survey goes some way towards investigating this by asking about previous use of other modes). Ultimately, the choice among different alternatives was made only on the basis of price changes, while attributes such as frequency of service, price of alternative modes, delay times and travel times were not mentioned to the respondent.

Another issue in this survey is respondents’ cognitive burden, as the hypothetical questions are mostly in iterative format. Ideally, a discrete choice experiment would substitute the content of the phone interview either with a lengthier computer administered survey on board or a follow-up post-phone survey. An experimental design could consider four alternatives: (a) current option of the respondent, (b) another coach service, (c) train and (d) car. Following the choice experiment, the derived discrete choice model could be used to develop alternative scenarios involving presence/absence of the different competitors.

**Table 18: Attributes and levels of the Stagecoach/Citylink 2006 case study**

Attribute	Alternative	Levels
Travel cost	Current option	Same as current, +1%, +10%, +30%, +50% of current option
	Another coach	Same as current, +1%, +10%, +30%, +50% of current option
	Train	Same as current, +1%, +10%, +30%, +50% of current option
	Car	Fixed - based on distance between origin and destination
Time of arrival	Current option	Ideal, -5 min, -15min, -30 min, -45 min of ideal time
	Another coach	Ideal, -5 min, -15min, -30 min, -45 min of ideal time
	Train	Ideal, -5 min, -15min, -30 min, -45 min of ideal time
	Car	Ideal, -5 min, -15min, -30 min, -45 min of ideal time

Finally, Table 19 provides a summary of the benefits and drawback of the hypothetical questions across the studies in accordance to methodological assessment presented in the previous chapters.

**Table 19: Qualitative assessment of the individual studies**

Study	For	Against
<b>1. Eastman Kodak Company &amp; Colour Care Limited (2001)</b>	<ul style="list-style-type: none"> <li>• Consumers choose among alternative service/product bundles. Choice exercise replicates consumer behaviour</li> <li>• Relatively simple cognitive task</li> <li>• Methodology allows consumer not to choose</li> <li>• Avoids ‘yea-saying’</li> <li>• Approach minimises non response and avoids outliers</li> </ul>	<ul style="list-style-type: none"> <li>• Prices of three attributes (ie next day, 3-day, and 6-day) increase at the same time in scenario, and therefore are correlated</li> <li>• True effects of each attribute increases from the base price to +5% or +10% cannot be recovered. Alternatively, a small orthogonal experimental design would have allowed control for</li> </ul>

Study	For	Against
	<ul style="list-style-type: none"> <li>Realistic view of the process as it is unlikely that same-day or mail-order customer would switch to a more expensive or increased price option</li> </ul>	<ul style="list-style-type: none"> <li>correlations among the prices and recover main effects</li> <li>There might be an anchoring bias ie consumer may wish to stick with the baseline preference</li> </ul>
<b>2. First Group Plc / ScotRail Railways Limited (2003)</b>	<ul style="list-style-type: none"> <li>Straightforward questions. Little cognitive burden to the respondent</li> <li>Low risk for anchoring or starting point bias as it does not provide respondents with cues about what the value of frequency, price, journey time might be</li> <li>Elicit directly the minima and maxima regarding the aforementioned attributes for each respondent</li> </ul>	<ul style="list-style-type: none"> <li>High risk of collecting outliers, non-response, zero answers and unreliable responses. Respondents, for example, were more tolerant to delays, number of trains to reach their destination than their present situation</li> <li>Respondents may be unfamiliar with the new alternative on offer and its characteristics (ie switching their journey from train to bus and <i>vice versa</i>)</li> <li>Questions may be interpreted as respondents' tolerance to deterioration of the service</li> <li>May fail to achieve enough variability required for the analysis</li> </ul>
<b>3. National Express Group PLC/ Greater Anglia Franchise (2004): Mix of dichotomous choice (yes/no) followed by open-ended questions</b>	<ul style="list-style-type: none"> <li>Respondents may think about only a single price to decide whether they would switch between modes (bus over train)</li> <li>Avoids non-response and outliers (in the dichotomous choice)</li> </ul>	<ul style="list-style-type: none"> <li>Respondents might not be aware of the competing coach service and the details about it</li> <li>High risk of 'yea-saying'; <i>"...is there a reduction in that train fare that would have been just enough to make you use the train instead of the coach?"</i> (No/Yes)</li> <li>Questions do not ask for maximum increase or reduction that would make respondents switch</li> </ul>
<b>3. National Express Group PLC/ Greater Anglia Franchise (2004): Discrete choice experiment</b>	<ul style="list-style-type: none"> <li>Realistic alternatives (eg coach is always slower than train)</li> </ul>	<ul style="list-style-type: none"> <li>Missing diagnostic questions, ie whether respondents understood the options given and made an informed choice</li> <li>Some choice exercises provided respondents with "easy" options</li> </ul>

Study	For	Against
<b>4. Acquisition of the Greater Western Passenger Rail Franchise (2005)</b>	<ul style="list-style-type: none"> <li>• Straightforward questions</li> </ul>	<ul style="list-style-type: none"> <li>• Opinion questions (increase, decrease of price, etc.) run the risk of respondents expressing a general dissatisfaction (protest) about prices in general</li> <li>• Intended behaviour question implies certain action will be surely taken by respondents</li> <li>• Respondents may be unfamiliar with all alternative options and their costs and, thus, may not be 'fully informed' as is implied by the models. A 'training' question (revealed preferences) might have been used to introduce respondents to alternative options and possible costs. Respondents could then be asked about their intended behaviour.</li> </ul>
<b>5. Stagecoach / Scottish Citylink (2006)</b>	<ul style="list-style-type: none"> <li>• Iterative bidding involving dichotomous questions help to identify the sensitivity of respondents to a range of prices</li> <li>• Help to identify more than one observation points from each respondent</li> <li>• May facilitate respondents' thought processes and encourage them to consider their preferences carefully</li> </ul>	<ul style="list-style-type: none"> <li>• Possible loss of incentive compatibility (truth telling) since the second, third, etc. question might be viewed by respondents as being exogenous to the choice situation, and the added possibility of 'yea-saying' and fatigue biases</li> <li>• There is risk of starting point bias since answers are 'anchored' on the initial figure stated by the questioner</li> <li>• Respondents may not be aware of the other available alternatives and their specific attributes (ie, price). A 'revealed preference' card with all available modes and prices would help to 'fully inform' respondents so that they can provide more realistic answers to the hypothetical questions</li> </ul>

### 5.3 Moving from CVM to DCE: An illustration using the Eastman Kodak Company (2001) study

#### Background

The example in this section is designed to illustrate a DCE equivalent to the original CVM Eastman Kodak Company (2001) study.

#### Introduction

The aim of the Eastman Kodak Company (2001) case study was to understand the behaviour of consumers with respect to film development and processing (D&P) services. In particular, the objective was to understand the likely behaviour of consumers in response to an increase in the price of next-day-to-within a week D&P.

The case study asked respondents about their choice of D&P time for 25-exposure APS film given the cost for different options. The study asked respondents to provide their choices under three scenarios/questions:

- Base case: offering current prices for same day, next day, 3-day, 6-day and mail-order service,
- Scenario 1: the price of next day, 3-day, and 6-day increase by 76p, and
- Scenario 2: There is a further price increase of next day, 3-day, and 6-day increase by 74p.

## Stated preference study design

Table 20 provides a task schedule and timescale for a stated preference survey based on recruiting 2,000 respondents aged 16 years or older in Great Britain. The total time to undertake the SP study ranges from 12 to 17 weeks, depending on the survey methodology adopted.

**Table 20: Tasks and timeframes for the Eastman Kodak Company study using a DCE**

Task	Details	Time required (days)		
<b>Background research</b>	<ul style="list-style-type: none"> <li>• What is the pricing inquiry being asked?</li> <li>• What is the product or service being valued?</li> </ul>	2		
<b>Choice of survey method</b>	<ul style="list-style-type: none"> <li>• Face-to-face, mail, internet, mix of methods</li> </ul>	2		
<b>Sampling of population</b>	<ul style="list-style-type: none"> <li>• Target population, sampling and method of recruitment</li> </ul>	3		
<b>Questionnaire design</b>	<ul style="list-style-type: none"> <li>• Coding of background questions</li> <li>• Development and coding of the stated choice experiment</li> <li>• Coding of follow-up questions</li> <li>• Sign-off from the CC</li> </ul>	10		
<b>Testing of questionnaire</b>	<ul style="list-style-type: none"> <li>• Pilot/pre-test survey(s)</li> </ul>	4		
<b>Analysis of pilot data</b>	<ul style="list-style-type: none"> <li>• Preliminary choice models</li> <li>• Re-design</li> <li>• Sign-off from the CC</li> </ul>	5		
<b>Main survey</b>	<ul style="list-style-type: none"> <li>• Conduct the main survey for 2,000 adults <ul style="list-style-type: none"> <li>– Face-to-face (F-t-F)</li> <li>– Post</li> <li>– Phone-post-phone (P-p-P)</li> <li>– Internet</li> </ul> </li> </ul>	15-30 10-20 15-30 5-10		
<b>Econometric Analysis</b>	<ul style="list-style-type: none"> <li>• Coding of database</li> <li>• Development of choice models</li> <li>• Quality assurance</li> </ul>	10		
<b>Validity and reliability testing</b>	<ul style="list-style-type: none"> <li>• Do the results meet validity and reliability tests?</li> </ul>	5		
<b>Development of scenarios</b>	<ul style="list-style-type: none"> <li>• Decision support system</li> <li>• Estimation of market shares for different changes at price levels</li> </ul>	5 3		
<b>Reporting</b>		4		
<b>Total time required</b>				
	<b>F-t-F</b>	<b>Post</b>	<b>P-p-P</b>	<b>Internet</b>
<b>Min (weeks)</b>	14	12	14	11
<b>Max (weeks)</b>	17	14	17	12

The first step in the SP study is to conduct background research in order to define the objectives of the study and the structure of the DCE experiment. The CC and analysts should agree, for example, on the different options available, price range (or changes) that would be applied for each of the processing and development options (same day, 3-day processing etc.), etc. While current options include same day, next day, 3-day service, 6-day service and mail-order, the study might consider, for example, differentiating between express-mail and regular-mail orders with different price changes for each. These price-changes would also have to be defined. Moreover, the CC and analysts should agree upon the base prices and hypothetical price changes for each of the processing and development options. The ideal estimated time to complete the background research stage is estimated at two days<sup>20</sup>.

An important decision is to decide upon the choice of the survey method. As mentioned in Appendix B the choice of the survey method depends largely on the cost and time resources that are available.

Interviewer administered interviews (ie face-to-face or telephone) are considered the most reliable option by SP practitioners as the interviewers can monitor responses to the SP tasks and assist if necessary. On the other hand, interviewer administered interviews are more expensive than self completion methods (Internet and postal). Postal methods do not allow for a design customised on earlier responses and are therefore less flexible than computer assisted methods. Online methods are increasingly being used for stated preference.

All methods except postal can be effectively used for SP surveys and the method chosen should be taken based on the sampling requirements, survey costs and timescale.

Finally, it is worth mentioning that regardless of the level of complexity, it would be very challenging to include an SP exercise in a meaningful way as part of an Omnibus survey or any other large scale data collection campaign. An SP study should be designed to answer the specific research question defined during the background research stage. The estimated time for deciding on the survey method is approximately two days.

The nature of the target population will partially dictate the nature of the sampling method. For example, a geographically spread sample can be effectively sampled by telephone or internet; high street shoppers can be effectively sampled in high streets, etc.

For all methods it is important to ensure that the sample is representative of the target population. This is typically undertaken through the use of quotas. The sampling procedure may take a minimum of three days depending on the nature of the sample.

The design of the questionnaire involves the write-up and coding of the background and follow-up questions and the design and coding of the choice experiment. The following section provides a detailed description of the experimental design in the context of the 2001 Eastman Kodak Company case study. The anticipated time to complete the design and coding of the questionnaire is a minimum of 10 days. This timeframe includes the

---

<sup>20</sup> At the point, we do not propose focus group research as we assume that we understand the issues around the research question well enough.

time required for the CC to sign off the questionnaire. Finally, it should be mentioned that more time would need to be allocated to the design of the survey if the experiment is more complex than the 2001 Eastman Kodak Company study.

## Experimental design and choice sets

This section demonstrates how the CVM scenarios could be developed using a DCE. Specifically, the series of three pricing scenarios/questions in the 25-APS film scenario could be substituted with a total of nine scenarios of a fractional-factorial design (see Figure 15). For each processing time on offer, the price varies only in the next day, 3-day, and 6-day options. Effectively, price is the only attribute varying across three alternative bundles. Therefore, using the  $L^{MA}$  design methodology, this would result into 27 ( $3^{1 \times 3}$ ) scenarios of a full factorial design. Further, we can reduce the number of scenarios further down to nine scenarios using a fractional factorial design, which would allow estimation of all main effects of the attributes. This procedure can be done using the design functionality within a number of different statistical packages (eg SPSS).

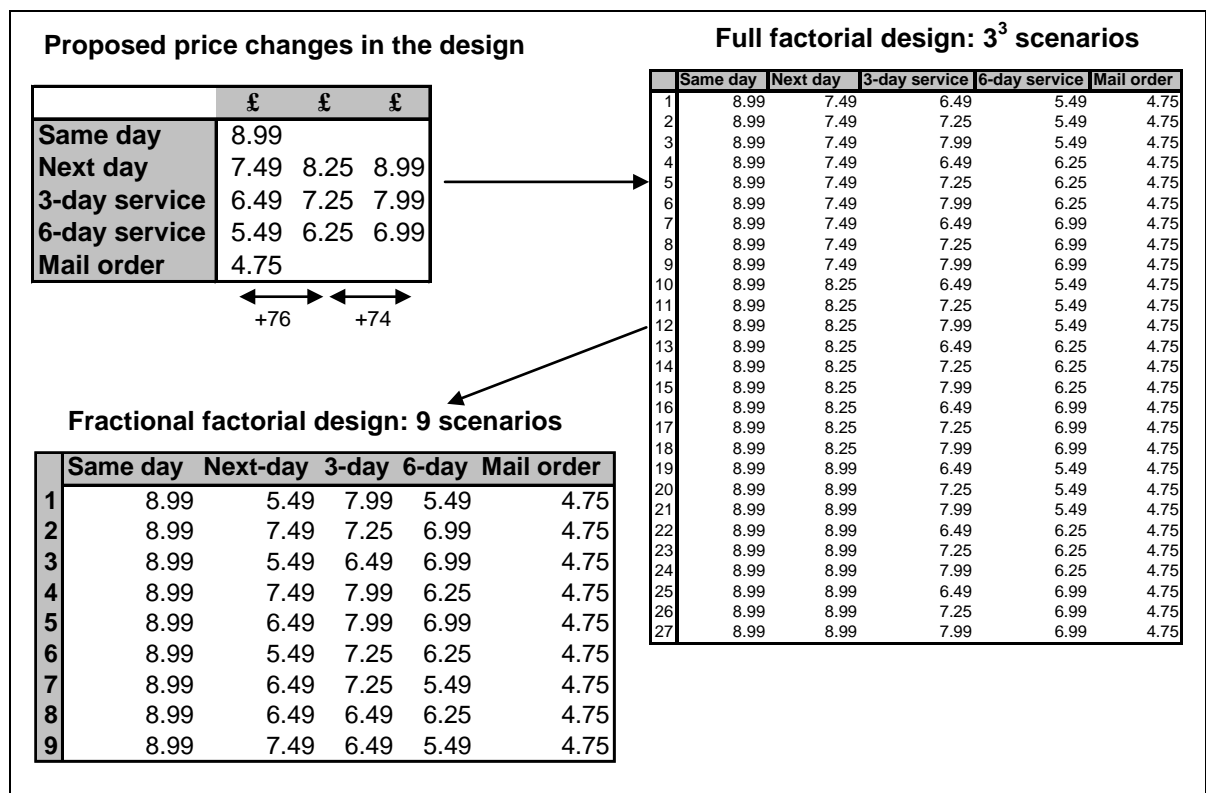


Figure 14: Experimental design of scenarios in the 25-APS film case-study<sup>21</sup>

## Choice task

Given that the choice task is relatively simple, each respondent could provide responses to all nine of the scenarios in the design. The format of the choice exercise for scenario No. 5 above would be as follows:

<sup>21</sup> Following testing or during the design phase, some scenarios where a quicker service is cheaper may be avoided.

*“Now, I would like you to focus on the following card, which gives the prices of different times of developing and processing services of a 25-APS film. Under this situation, which of the following processing times would you choose?”*

	<b>Same day</b>	<b>Next-day</b>	<b>3-day</b>	<b>6-day</b>	<b>Mail order</b>
	<b>£8.99</b>	<b>£6.49</b>	<b>£7.99</b>	<b>£6.99</b>	<b>£4.75</b>
<b>I would choose</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The benefits of using the DCE approach over the CVM questions are the following:

- Each respondent provides more data points/observations (ie nine), so more information is available for model development
- Model results are more efficient ie the analyst is more certain about the influence of price change in the demand for different processing times
- Direct and cross-elasticities of different product/price bundles can be estimated
- A decision support system using the developed models can provide more reliable market shares of the different processing times on offer.

As mentioned in Appendix B testing the design and assessing the quality of responses is an essential part in any type of the survey. Testing the SP survey and the quality of the collected data becomes even more important as it helps control or avoid a number of biases listed in Table 12. Failure to address problems at the pilot stage can lead to meaningless results at the end of the study. An important element in the analysis of the pilot and the main data is the organisation of the data sets. Different software requires the data in different forms. Quality control procedures should be used to ensure that incomplete observations, outliers and incorrect entries have been removed. Therefore, a minimum of nine days should be allocated for testing and conducting preliminary analysis of the pilot data.

The practice of CC not to allocate time for pilot testing is understandable in the context of relatively simple CVM questions, for example as part of an Omnibus survey. However, this does not reflect the recommended practice for SP surveys. A final sign-off by the CC would be necessary at the end of both the questionnaire design and the analysis of the pilot data stages. The total estimated time for designing the survey, testing and analysis of the pilot data is 19 days.

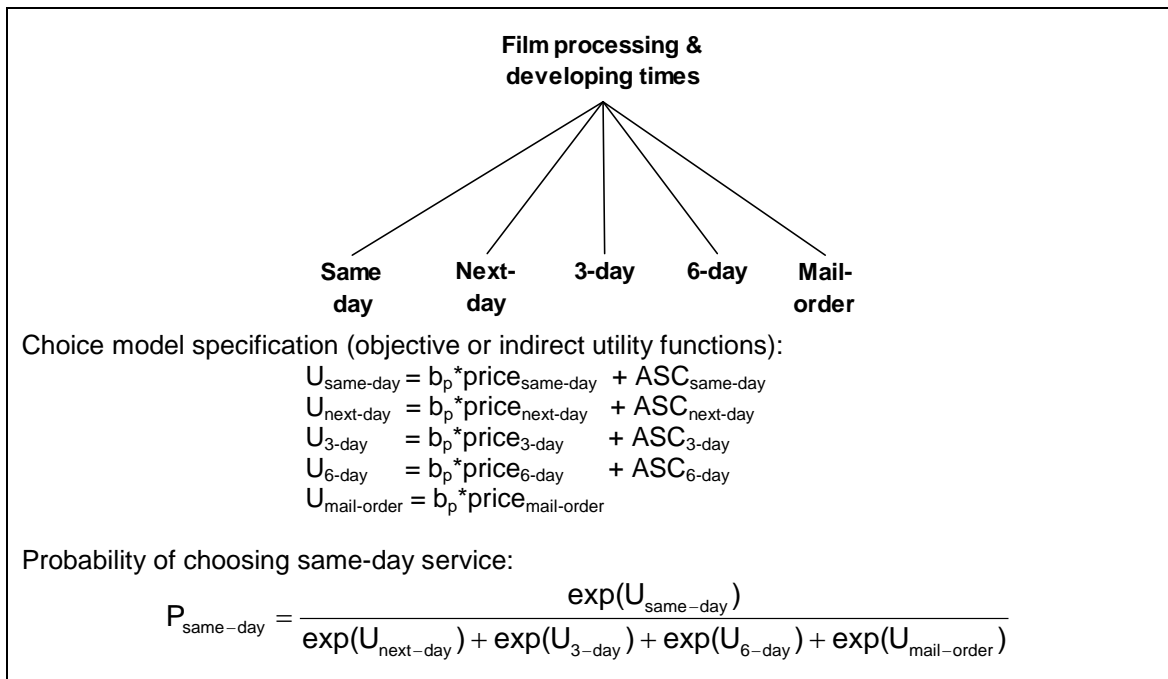
### **Main survey and econometric analysis**

The time required to complete the main survey is dependent on the type of survey methodology adopted. We have allocated 15-30 days for face-to-face interviews, 10-20 days for postal survey, 15-30 days for phone-post-phone and 5-10 days for an Internet survey.

In this case study, the objective of the analysis is to establish market shares in the sample and to determine whether individuals would switch from one film processing and developing option to another given different price changes. Therefore, the first step is to develop discrete choice models based on random utility theory. Figure 16 shows the structure and formulation of a conditional multimomial-logit (MNL) choice model,

which the most widely used model structure in discrete choice analysis (Ben-Akiva and Lerman, 1985).

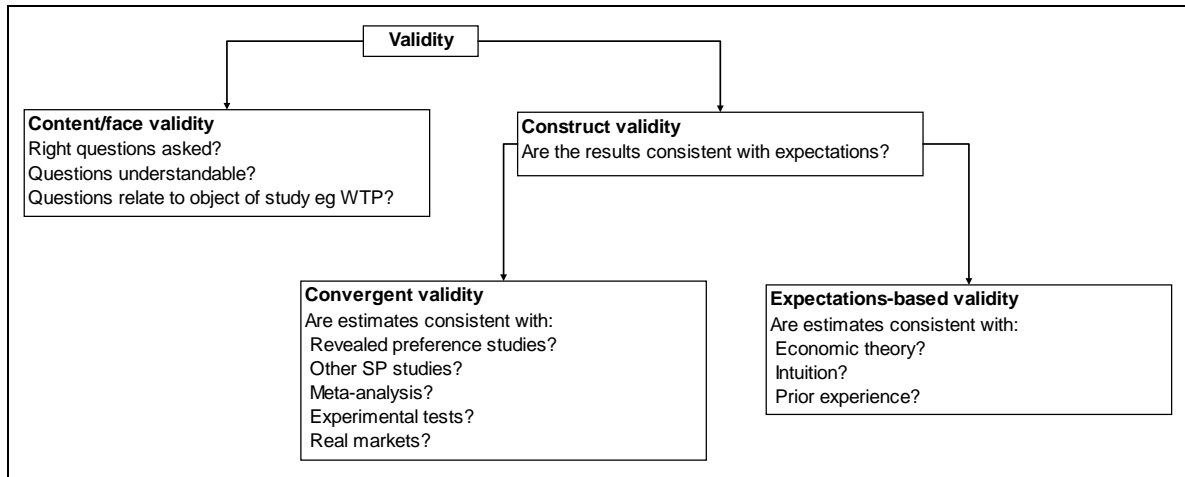
The model formulation in Figure 16 assumes that price is the only characteristic that influences choice of different film processing and developing services. Estimation of the model using the collected DCE data will enable the analyst to express the change in utility from the hypothetical change in price. Further analysis may involve segmentation variables such as socio-economic and demographic characteristics of individuals. The analysis of the choice data becomes more complex with additional attributes. The minimum time for data coding, developing the choice models and evaluating the outcome through a quality assurance process is estimated at 10 days.



**Figure 15: Structure of the choice model**

### Validity and reliability testing

Tests of validity are applied to examine whether the biases discussed in Table 12 exist. Validity refers to the degree to which a study is successful in measuring individuals' preferences by overcoming potential biases and the hypothetical nature of choice exercise (Pearce and Ozdemiroglu, 2002). Reliability refers to the degree of replicability of a measurement. Reliability exercises typically involve the repetition of the case study at different points in time and so are considered a reasonable requirement for each individual study. The different types of validity testing are described in Figure 17. The estimated time to conduct different validity tests is five days.



**Figure 16: Types of validity tests**

## Development of scenarios

Following the validity testing, the utility expressions estimated using the choice data can be used to simulate different scenarios<sup>22</sup> of price changes among the different development and processing services. Specifically, the model can be applied to gain insight into the likely volume shares across D&P services under different price scenarios. For example, one scenario might examine the likely change in market share when the 3-day and 6-day service prices increase by 50p. The decision support system for simulating scenarios can be in the form of an Excel application. The estimated time to develop the decision support system and run the scenarios is eight days.

## Reporting

Reporting is the last stage of the SP case study. This provides a transparent and comprehensive presentation of the approach followed and the results of the study. The main elements of the report should include:

- Executive summary
- Objectives of the study
- Review of any previous relevant studies
- Description of the survey design
- Description of the survey administration procedure
- Summary of data
- Details of the data analysis
- Tests of validity
- Results obtained from simulation scenarios, and
- A copy of the questionnaire.

The estimated time for reporting is four days.

<sup>22</sup> Prior to this stage, the analyst would need to collect information on the current service used and the perceived costs of all of the alternatives. Moreover, it would be necessary to calibrate the model to reproduce the base shares.

## 5.4 Recommendations from Review of Current Competition Commission surveys

As shown from the practical assessment of CVM and DCE in Table 15 and given that the CC studies have to be delivered within short timeframes, it is inevitable that the CVM type of questions is often the preferred option. However, on the basis of our review of a subset of previous Competition Commission surveys we would recommend consideration of the following points when undertaking future surveys:

- The analyst should explicitly consider the extent to which they wish to explore the respondents' current understanding of the choices that they have available, the extent to which any switching may be driven by changes in price (or service), and whether there is a risk that the survey instrument may confound the response to a price signal and the response to an improved understanding of the choices available that have been spelt out in the framing of the survey question. The latter issue is particularly pertinent when presenting respondents with closed form questions which provide a series of pre-defined response options which the respondent may or may not have perceived as viable, or considered as available alternatives, prior to being posed the question in this manner. Clearly, a confounding of the response to the price signal and the response to the improved information can result in a potential over-estimate of the price elasticity. In cases where the analyst wishes to unpick these issues it can be informative to include an additional stage within the questioning, where the respondent is informed of the current prices of possible alternatives and given an opportunity to indicate a change in behaviour in response to this improved information prior to varying the prices.
- The analyst should also take care to ensure that the question being asked is the question intended ie should remove any ambiguity that could lead to at best increased noise in the responses and at worst answers to an entirely different question to that envisaged. For example, in many cases when framing a contingent valuation question the analyst may be seeking to elicit the maximum willingness to pay or the minimum willingness to accept a stated change in service. However, the omission of the words "maximum" or "minimum" can introduce significant ambiguity in to the task which may not be appreciated by the analyst who will know what they intended to ask rather than possibly focusing upon what they have asked.
- In some of the examples examined there are cases where the survey sought to examine situations where attributes may be varying across a number of alternatives (eg the prices of the different processing options in the Eastman Kodak Company example). In such cases, the analyst may wish to consider adopting a simple experimental design, which allows the pricing scenarios being presented to be defined in a manner that will allow the impact of the prices on each of the presented alternatives to be understood independently from each other. In some of the SSNIP tests observed the decision has been taken to apply simultaneous price increases across multiple products. Whilst a set of these questions allows an understanding of the relevant market, the simultaneous increases in price do not allow an understanding of which product within the bundle is defining the market. In contrast, a more structured approach utilising an experimental design could still provide an understanding of the price response whilst allowing it to be further unpicked to explore the differential response to price increases across the different products within the bundle.

- The analyst can also benefit from considering the wording used to introduce the choice or valuation task. It is useful to both set the context for the issue under consideration (ie narrow in on the specific context) whilst reminding the respondent that this falls within their wider day-to-day considerations (eg time or budget constraints). The choice of wording can also have an impact on the way that the respondent approaches the task and the likelihood of protest responses that can bias the findings. For example, it may be better to avoid wording such as “in your view ...” which invites a value based judgement and rather use wording such as “imagine that you face the situation where ...” which places the respondent within the context of a more routine choice situation.
- There is also benefit from including diagnostic questions at the end of a survey to explore the extent to which the respondents understood any tasks that they had been asked to undertake or to give an opportunity to voice issues which they wanted to raise but felt that the survey questions did not give them opportunity to express. Such data can be useful in cleaning a data set and identifying those that may not have fully understood the questions. This form of follow-up question can also be used after contingent valuation tasks to explore the rationale for particularly low or high values to gain a better understanding of whether the response is a protest vote or a true high or low value.

## 5.5 Market Research Methodology Recommendations

### Introduction

This section focuses on our recommendations for the market research aspects of CC studies.

### Ways of dealing with bias

It is important to minimise the following forms of bias when selecting the sample:

- **sampling error** – the sample not being representative of the population as a whole. Sampling error is minimised by ensuring the sample is randomly selected but this is not always practical for reasons of time, cost and the nature of the sample. We recommend that the CC uses quota samples for their surveys whilst making every effort to ensure that the sampling approach is as random as possible.
- **non-response error** – those not responding being different from those that do. Non response error can be minimised by maximising response rates. We recommend that this is done through the following:
  - using short well designed questionnaires
  - using well trained interviewing staff
  - giving assurances of confidentiality and data protection
  - mentioning the survey sponsor’s name
  - using of incentives (if the questionnaire is long and/or complex).

Face to face methods tend to have the best response rates, followed by telephone methods with self completion methods, particularly postal, having the worst response rates.

## **Cheap talk**

We recommend that the ‘cheap talk’ approach is considered to correct for hypothetical bias in DCE and CVM surveys in particular. The idea behind the ‘cheap talk’ approach is that a *cheap talk script* makes respondents aware of the problem of hypothetical bias prior to the actual hypothetical questions. The advantage of the approach is that it provides an *ex ante* bias correction that can be applied to any evaluation.

## **Use of CC's name**

We recommend that CC’s name is mentioned in their surveys as the survey sponsor to increase the response rate. We do not consider that mentioning the CC as a survey sponsor will cause any bias

## **Length of questionnaire**

It is important to ensure that the questionnaire is not too long as longer questionnaires tend to lower responses rates and make the survey less representative as well as increasing costs. Incentives can help alleviate concerns with respect to response rates for longer interviews.

The recommended maximum length of the question is partly dependent on research methodology with, for example, longer questionnaires being possible for in home interviews and shorter questionnaires required for *in situ* interviews.

It should also be noted that DCE questionnaires are typically at least 20 minutes in length.

We recommend that CC questionnaires are restricted to 10 minutes in duration where practical and are no longer than 30 minutes even if using DCE.

## **Marginal customers**

Our recommendation with respect to identifying marginal customers is to undertake qualitative research to help identify the nature of such respondents and to help define the appropriate question or questions needed for such an identification.

Our recommendations on scaling marginal customers’ responses to the overall population is to ensure that non marginal customers are also be sampled in the research and to ensure that the proportion of marginal customers to all customers is correctly identified so that weighting of the data can be undertaken.

## **Sample size**

As a larger sample size leads to more robust data there is always a trade off in market research between cost and precision. In this respect, statistical reliability is a relative term. We recommend using the market research common practice of using a minimum total sample size of 400 for simple surveys as this represents a reasonable balance between robustness of results and cost of fieldwork.

However, it need to be noted that the sample size for each specific sub group of interest needs to be large enough to provide for robust data taking account of the statistical significance when comparing data between groups.

Finally, for stated preference surveys it is recommended that the size of each cell of interest is at least 75.

## **Pilot**

The pilot is an ideal that should be used whenever possible as part of the research process, particularly for stated preference approaches. It is used to test both the survey methodology and the questionnaire.

The pilot sample size for stated preference research should be at least 50.

If time constraints mean a full pilot is not possible it is recommended that the main survey is undertaken for a day or two followed by a break of a couple of days for analysis and review if possible.

## **Out of Scopes/Refusals**

We recommend that for all surveys the number of people screened and found to be out of scope and the number who refuse to take part are recorded, so that an estimate of the size of the overall population can be made and to establish whether refusals are representative of the sampled population.

-----

## 6. REFERENCES

- Adamowicz, V. & Boxall, P. (2001) Future directions of stated choice methods for environment valuation. *Choice experiments: A new approach to environmental valuation*. London, UK, April 10.
- Adamowicz, W., Boxall, P., Williams, M. & Louviere, J. (1998a) Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *Journal of Agricultural Economics*, 80, 1, 64-75.
- Adamowicz, W., Louviere, J. & Williams, M. (1998b) Introduction to attribute-based stated choice models. *US Department of Commerce, NOAA*.
- Amaya-Amaya, M., Gerard, K. & Ryan, M. (2008) Discrete choice experiments in a nutshell. In Ryan, M., Gerard, K. & Amaya-Amaya, M. (Eds.) *Using discrete choice experiments to value health and health care*. Springer, Dordrecht.
- Amelio, A. & Donath, D. (2009) Market definition in recent EC merger investigations: The role of empirical analysis. *Concurrences*, 3, 1-6.
- Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. & Shuman, H. (1993) Report of the NOAA panel on contingent valuation. *Federal Register*, 58, 4601-4614.
- Bateman, I., Munro, A., Rhodes, B., Starmer, C. & Sugden, R. (1997) A theory of reference-dependent preferences. *Quarterly Journal of Economics*, 112, 479-505.
- Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozedemiroglu, E., Pearce, D., Sugden, J. & Swanson, J. (2002) *Economic valuation with stated preference techniques: A manual*, Edward Elgar, Cheltenham, UK.
- Bateman, I.J., Lovett, A.A. & Brainard, S.J. (2005) *Applied environmental economics: A GIS approach to cost-benefit analysis*, Cambridge University Press, Cambridge.
- Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. & Rao, V. (1994) Combining revealed and stated preferences data. *Marketing Letters*, 5, 4, 335-350.
- Ben-Akiva, M. & Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge.
- Bennett, J. & Blamey, R.K. (2001) *The choice modelling approach to environmental valuation*, Edward Elgar Publishing Limited, Cheltenham.
- Bliemer, M. & Rose, J. (2006) Designing stated choice experiments: The state of the art. *11th International Conference on Travel Behaviour Research*. Kyoto, Japan.
- Bliemer, M.C. & Rose, J. (2005) *Efficiency and sample size requirements for stated choice studies*, Institute of Transport and Logistic Studies, The University of Sydney, Sydney.
- Boxall, P., Adamowicz, V., Swait, J., Williams, M. & Louviere, J. (1996) A comparison of stated preference methods for environmental valuation. *Ecological Economics*, 18, 243-253.

- Bradley, M.A. & Daly, A.J. (1997) Estimation of logit choice models using mixed stated preference and revealed preference information. In Stopher P. & Lee-Gosselin, M. (Eds.) *Understanding Travel Behaviour in an Era of Change*. Elsevier, Oxford
- Bruce, C. (2006) Can contingent valuation resolve the "adding-up problem" in environmental impact assessment? *Environmental Impact Assessment Review*, 26, 570-585.
- Burge, P., Devlin, N., Appleby, J., Rohr, C. & Grant, J. (2005) London Patient Choice Project Evaluation. RAND Europe, Cambridge.
- Cameron, T.A., Shaw, W.D., Ragland, S.E., Callaway, J.M. & Keefe, S. (1996) Using actual and contingent behavior data with differing levels of time aggregation to model recreation demand. *Journal of Agricultural and Resource Economics*, 21, 130-149.
- Carlsson, F., Frykblom, P. & Lancaster, K.J. (2004) *Using Cheap-Talk as a Test of Validity in Choice Experiments*, Department of Economics, Gothenburg University. Working Papers in Economics no. 128.
- Carlsson, F. & Martinsson, P. (2003) Design techniques for stated preference methods in health economics. *Health Economics*, 12, 281-294.
- Carson, R.T., Flores, N.E., Martin, K.M. & Wright, J.L. (1996) Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. *Land Economics*, 72, 1, 80-99.
- Carson, R.T., Louviere, J., Anderson, D.A., Arabie, P., Bunch, D.S., Hensher, D.A., Johnson, R.M., Kuhfeld, W., Steinberg, D., Swait, J., Timmermans, H. & Wiley, J.B. (1994) Experimental analysis of choice. *Marketing Letters*, 5, 4, 351-368.
- Cave, M., Burningham, D., Buxton, M., Hanney, S., Pollit, C., Scanlan, M. & Shrumer, M. (1993) The valuation of changes in quality in the public services. HMSO, London.
- Champ, P.A. & Welsh, M.P. (2006) Survey methodologies for stated choice studies. In Kanninen, B.J. (Ed.) *Valuing Environmental Amenities Using Stated Choice Studies*. Springer, Heidelberg.
- Cherchi, E. & Ortázar, J.D.D. (2002) Mixed RP/SP models incorporating interaction effects: Modelling new suburban train services in Cagliari. *Transportation*, 29, 4, 371-395.
- Cummings, R.G., Schulze, W.D., Gerking, S. & Brookshire, D.S. (1986) Measuring the elasticity of substitution of wages for municipal infrastructure: A comparison of the survey and wage hedonic approaches. *Journal of Environmental Economics and Management*, 13, 269-276.
- Cummings, R.G. & Taylor, L.O. (1999) Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *The American Economic Review*, 89, 3, 649-665.
- Englin, J. & Cameron, T.A. (1996) Augmenting travel cost models with contingent behavior data. *Environmental and Resource Economics*, 7, 133-147.
- Finn, A. & Louviere, J.J. (1992) Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy & Marketing*, 11, 1, 12-25.

- Flynn, T.N., Louviere, J.J., Peters, T.J. & Coast, J. (2007) Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, 26, 171-189.
- Foster, V. & Mourato, S. (2003) Elicitation format and sensitivity to scope - Do contingent valuation and choice experiments give the same results? *Environmental & Resource Economics*, 24, 2, 141-160.
- Fox, J., Shogren, J., Hayes, D. & Kleibenstein, J. (1999) CVM-X: Calibrating contingent values with experimental auction markets. *American Journal of Agricultural Economics*, 80, 3, 455-465.
- Freeman, M.A.I. (2003) *Environmental and Resource Values: Theory and Methods*, Resources for the Future Press, Washington D.C.
- Green, P. & Rao, V. (1971) Conjoint measurement for quantifying judgement data. *Journal of Marketing Research*, 8, 355-363.
- Green, P.E. & Srinivasan, V. (1978) Conjoint Analysis in Consumer Research - Issues and Outlook 2. *Journal of consumer research*, 5, 2, 103-123.
- Hanemann, M. (1984) Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, 66, 3, 332-341.
- Hanley, N., Wright, R.E. & Adamowicz, V. (1998) Using choice experiments to value the environment - Design issues, current experience and future prospects. *Environmental and Resource Economics*, 11, 3-4, 413-428.
- Hensher, D. & Bradley, M. (1993) Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters*, 4, 2, 139-151.
- Hensher, D.A. & Greene, W.H. (2003) The Mixed Logit model: The state of practice. *Transportation*, 30, 133-176.
- Hensher, D.A., Rose, J.M. & Greene, W.H. (2005) *Applied Choice Analysis - A Primer*, Cambridge University Press, New York.
- Hess, S., Rose, J. & Polak, J. (2008) Non-trading lexicographic and inconsistent behaviour in SP choice data. *87th Annual Meeting of the Transportation Research Board*. January 13-17, Washington D.C.
- Hidona, N. (2002) *The Economic Valuation of the Environment and Public Policy - A Hedonic Approach*, Edward Elgar Publishing Limited, Cheltenham, UK.
- Huber, J. & Swerina, K. (1996) The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33, 3, 307-317.
- Johnson, F.R., Kanninen, B., Bingham, M. & Ozdemir, S. (2006) Experimental design for stated choice studies. In Kanninen, B.J. (Ed.) *Valuing environmental amenities using stated choice studies*. Springer, Dordrecht.
- Kahneman, D., Knetsch, J.L. & Thaler, R.H. (1990) Experimental tests of the endowment effect and the coarse theorem. *Journal of Political Economy*, 98, 6, 728-741.
- Keane, M. (1997) Current issues in discrete choice modelling. *Marketing Letters*, 8, 3, 307-322.
- Kjaer, T. (2005) A review of the discrete choice experiment - with emphasis on its application in health care. Health Economics, University of Southern Denmark.

- Kroes, E.P. & Sheldon, R.J. (1988) Stated preferences methods: An introduction. *Journal of Transport Economics and Policy*, 22, 1, 11-25.
- Kuhfeld, W. (2009) *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint and Graphical Techniques*, SAS Institute Inc., Cary, NC, USA.
- Lancaster, K.J. (1966) A new approach to consumer theory. *Journal of Political Economy*, 74, 2, 132-157.
- List, J.A. & Gallet, C.A. (2001) What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental & Resource Economics*, 20, 3, 241-254.
- Loomis, J., Gonzalez-Caban, A. & Gregory, R. (1994) Substitutes and budget constraints in contingent valuation. *Land Economics*, 70, 4, 499-506.
- Louviere, J. & Lancsar, E. (2009) Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy and Law*, 4, 527-546.
- Louviere, J. & Woodworth, G. (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. *Journal of Marketing Research*, 20, 350-367.
- Louviere, J.J. (1988) Conjoint analysis modelling of stated preferences: A review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy*, 22, 93-119.
- Louviere, J.J. (2000) Why stated preference discrete choice modelling is not conjoint analysis (and what SPDCM is). *Memetrics White Paper*.
- Louviere, J.J., Hensher, D.A. & Swait, J.D. (2000) *Stated Choice Methods: Analysis and Application*, Cambridge Press, Cambridge.
- Lusk, J.L. (2003) Effects of cheap talk on consumer willingness-to-pay for golden rice. *American Journal of Agricultural Economics*, 85, 4, 840-856.
- McFadden, D. (1973) Conditional logit analysis of qualitative choice behaviour. In Zerebka, P. (Ed.) *Frontiers in Econometrics*. Academic Press, New York.
- Moresi, S.X., Salop, S.C. & Woodbury, J.R. (2008) Implementing the hypothetical monopolist SSNIP test with multi-product firms. *The Antitrust Source*, February, 1-8.
- Ortuzar, J.D. & Willumsen, L.G. (2001) *Modelling Transport*, John Wiley & Sons, Ltd, Chichester.
- Pearce, D. & Ozdemiroglu, E. (2002) *Economic valuation with stated preference techniques*, Department for Transport, Local Government and the Regions, Rotherham.
- Randal, A. (1994) A difficulty with the travel cost method. *Land Economics*, 70, 88-96.
- Ready, R.C., Buzby, J.C. & Hu, D. (1996) Differences between Continuous and Discrete Contingent Valuation Estimates. *Land Economics*, 72, 3, 397-411.
- Ryan, M. (1999) A role for conjoint analysis in technology assessment in health care? *International Journal of Technology Assessment in Health Care*, 15, 3, 443-457.

- Schulze, W., McClelland, G., Waldman, D. & Lazo, J. (1996) Sources of Bias in Contingent Valuation. In Bjornstad, D.J. & Kahn, J.R. (Eds.) *The Contingent Valuation of Environmental Resources*. Oak Ridge National Laboratory, U.S. Edward Elgar,
- Street, D.J., Burgess, L., Viney, R. & Louviere, J. (2008) Designing discrete choice experiments for health care. In Ryan, M., Gerard, K. & Amaya-Amaya, M. (Eds.) *Experiments to value health and health care*. Springer, Dordrecht.
- Train, K. (1993) *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, 3rd Edition, MIT Press, London.
- Von Haefen, R.F. & Phaneuf, D.J. (2008) Identifying demand parameters in the presence of unobservables: a combined revealed and stated preference approach. *Journal of Environmental Economics and Management* 56, 19-32.
- Wardman, M. (1988) A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy*, 22, 1, 71-91.
- Whitehead, J.C., Pattanayak, S.K., Van Houten, G.L. & Gelso, B.R. (2008) Combining revealed and stated preference data to estimate the nonmarket value of ecological services: An assessment of the state of science. *Journal of Economic Surveys*, 22, 5, 872-908.

# APPENDIX A

## International Depths Topic Guide



Interviewer: .....

Respondent: .....

Title and business name: .....

Date and time: .....

Good morning/afternoon. My name is ... and I work for an independent market research company called Accent. We have been commissioned by the UK Competition Commission to conduct research into the research methods used by competition commissions and authorities both in the UK and elsewhere.

Part of this study involves talking to representatives of competition commissions or authorities in other countries.

Thank you very much for agreeing to help us with this research. The research is being conducted in accordance with the Code of Conduct of the UK Market Research Society and also with the UK Data Protection Act, with whom Accent is registered.

Some of the questions may cover information which you would not wish to be publicised. If this is the case, please let me know and I assure you that we will treat any such information as confidential.

The length of the discussion will depend on your involvement in this work but will not be longer than 30 minutes.

**Ask all**

- What is the nature of evidence required on willingness to pay and demand elasticities for the different types of investigations that you may undertake?

**THIS MAY LEAD YOU TO REALISE THEY CAN'T TELL YOU ANYTHING AND THEREFORE END THE INTERVIEW OR THEY MAY POINT YOU TO ANOTHER COLLEAGUE IN THEIR ORGANISATION WHOM WOULD BE BETTER TO TALK TO.**

**Use of Research**

- To what extent do you commission market research as part on competition investigations? Can you tell me more about this?
  - IF YES, PROBE TYPE OF RESEARCH, TYPE OF COMPETITION ISSUES
  - IF NO, WHY NOT, HAVE YOU CONSIDERED? IF YES, WHY REJECTED

### Those who use market research

- How often do you commission market research as part of competition investigations? What type of questions do you use such research to explore?
- Are you using market research as part of competition investigations more or less than before, do you intend to use it in the future? Why?
- What are your requirements from the 'hypothetical' research questions?
- What research techniques do you use to examine customers' responses to changes in price or other product attributes **PROBE, IF NECESSARY, SNIP TESTS, TRANSFER PRICING, RESERVE PRICING, QUALITATIVE RESEARCH, STATED PREFERENCE, CONTINGENT VALUATION ETC**
- What do you see as the strengths and values of these techniques?
- Do you have any concerns with any of these techniques? **PROBE NATURE OF CONCERNS BY TECHNIQUE**

- **IF SP/CONTINGENT VALUATION TECHNIQUES USED.** What would you say are the strengths and weaknesses of SP/contingent valuation in relation to other techniques
  
- **IF SP/CONTINGENT VALUATION TECHNIQUES NOT USED.** Have you considered using SP or contingent valuation? **IF SO, WHY REJECTED**
  
- What would you say is the most effective research approach for price sensitivity questions? **PROBE TECHNIQUES, APPLICATION (IE COMPUTER/FACE TO FACE/SAMPLING ISSUES ETC)**

### **Communication of Results**

- How do you communicate results from SP to your decision makers? Do they accept the results? What are their concerns with results?

Those are all the questions I have. Is there anything else which you think would be relevant?

## APPENDIX B

### Other Survey Issues

## OTHER SURVEY ISSUES

### Market Research Method

Stated preference and willingness to pay studies can be complex and need to be designed with care. A sound design of the market research methodology is essential to ensure that the results of the survey are robust. The market research method (along with the stated preference or WTP design) is subject to detailed scrutiny in competition enquiries and poor survey design could undermine the findings.

This section outlines the steps that should be followed to ensure that the stated preference/WTP studies are methodologically sound.

A number of interview methods can be used including in home, on street, telephone and internet. We discuss when and where the different methods are appropriate and the advantages and disadvantages of each below.

The recommended market research methodology for stated preference and other WTP studies ideally involves a number of stages.

The recommended staged approach is shown below:



Each of these stages is described below.

### Qualitative Research/Workshop

For stated preference research in particular, if time and budget allows, the research greatly benefits from a workshop and a qualitative stage at the commencement of the study.

## Workshop

A workshop at the commencement of the study ensures that the consultancy or study designers fully understand the requirements from the research exercise. It provides a means of:

- ensuring any relevant pre-existing research and relevant data can be made use of
- confirming deliverables and expectations
- gaining a consensus on research issues
- discussing topics for inclusion in the research
- discussing in-scope respondents for the research
- agreement on the research process.

## Qualitative Research

For stated preference studies qualitative research can be used to:

- Broadly establish the attributes for the different areas under consideration and make sure existing attributes that have been previously identified are correct
- Identify any emerging attributes
- Understand how the different attributes can be assessed and whether this is logical for customers
- Test the language, description and images of attributes which will be included in the stated preference
- Explore willingness to pay levels so that appropriate upper and lower limits ('boundary' values) can be established.

The qualitative stage can include focus groups and/or depth interviews.

### *Focus groups*

A focus group is a discussion conducted by a moderator among a small group of respondents.

The key benefit of groups is that it facilitates discussion and the exchange of views across different customers, allowing for the sharing and challenging of views and understanding how people respond and feel about different attributes and levels as well as testing the language and definitions.

In terms of process, twelve respondents are normally recruited for each group with the aim of 6-8 attending (some allowance for drop out is always necessary). A standard group would be expected to last for one and a half hours.

In some cases an extended group will be held lasting for two hours. Typically these are held if the research includes a creativity session or a range of concepts are to be tested. They may also be used for *deliberative* groups where the first part of the group is used to ascertain views prior to any information being provided to respondents with the

second part testing reactions once the group has been informed about areas they may up until that stage been unaware of.

Further variations from the standard groups include mini-groups (of 4 or 5 participants), triads and paired depths. These tend to be used to explore more sensitive issues or specialist topics.

Groups can be held at viewing facilities or hotels. The former enables a number of client representatives to attend and view the first groups in order to be able to suggest any required or desired changes for the subsequent groups, which could then be held in hotels.

Recruitment and attendance levels can be enhanced by the offer of an incentive. This can take the form of a cash donation, a voucher from a large store or a donation to a charity of choice. The choice and size of incentive would vary according to the type of respondent.

### ***Depth interviews***

Typically depth interviews are best for testing the quantitative material whenever this is required as this is a very individual exercise. Also, depths are appropriate for customers who are difficult to recruit or when the issue to be discussed is sensitive.

The depths could be conducted in a variety of ways such as:

- **Qualitative Hall Clinics** – respondents are recruited directly off the street in a number of different and appropriate geographical areas. This is a cost effective methodology as, once respondents are recruited off street, they are taken into a dedicated location and participate in a depth interview.
- **Depths in home** – respondents are recruited directly off the street or by phone and interviewed at home by an executive
- **Telephone depths** – respondents are recruited by phone and an appointment is made for a follow up telephone interview by an executive. This method suffers by not being able to use show material unless posted or emailed out.

## **Design**

This stage includes the design of the stated preference/WTP questions. This is covered in detail in Chapter 3.

Other aspects of design include sampling, sample size and questionnaire design. Each of these is discussed below.

## **Sampling**

The stated preference or WTP questionnaire will be administered with a defined target population. The method of sampling can ensure that this sample is representative. However, it should be noted that often research is required into the behaviour of specific market segments, and with minimum sample requirements for each market segment, the

overall survey will not be representative of the population (and may require re-weighting).

The nature of the sample is dependent on the nature of the target population. This can include users of a particular product or service, residents of a particular area, employees of a particular business etc. It may also be important to compare views or attitudes with a control group (for example non users or users of a competing product or service, residents of another area etc).

It is important to minimise the following forms of bias when selecting the sample:

- sampling error – the sample not being representative of the population as a whole
- non-response error – those not responding being different from those that do.

### **Minimising sampling error**

Sampling error is minimised by ensuring the sample is randomly selected.

To ensure the sample is truly random a random probability sample is required. This ensure that every person has a known non zero probability of being sampled. There are many forms of probability sampling such as simple random, systematic, stratified, probability proportional to the number in the population and multistage.

The key benefit of probability sampling approaches is that the uncertainty in estimates based upon the data can be assessed.

The nature of CC enquiries means that the nature of the samples are likely to be specific target markets, such as users of a particular product or service. It is unlikely to be the general population and this means that, in practice, it will often be difficult to use a probability sampling approach because a comprehensive sample frame may not be readily available.

Most market research similarly deals with specific target groups and the methods employed are chosen on the grounds of practicality and cost. For example, if one wanted to sample users of a particular shop an effective sampling approach is to intercept shoppers at that shop. Similarly, rail passengers of a particular service can be effectively sampled on that rail service. It should be noted that such sampling approaches oversample frequent users of the service or product.

Telephone interviewing is widely used in market research as it is much more cost effective than face-to-face household interviewing and easily allows for a geographically spread sample. However, not everyone has a fixed-line phone and the sampling base (typically Random Digit Dialling sample) is not truly random.

To counter such biases quotas representing the target population are typically used. Quotas can be used to address the most obvious biases in different sampling approaches (for example towards frequent users of a service for intercept surveys, women and older respondents in telephone interviews, non workers in on street surveys). Quotas can also be set to represent the background population, for example to age and gender from the Census, although it may be more difficult to assess the appropriate characteristics of specific target groups, for example visitors to an amenity.

Whilst quota samples do not (strictly) allow the same inferences to be drawn as do random samples, confidence intervals are routinely calculated for them using the same statistical theory. These intervals usually provide a reasonable measure of the uncertainty inherent in the sampling process.<sup>23</sup>

In conclusion, we advocate that the CC does use quota samples for their surveys whilst making every effort to ensure that the sampling approach is as random as possible.

### **Minimising non response error**

Non response error can be minimised by maximising response rates. Ways to help maximise response rates include:

- Short well designed questionnaires
- Well trained interviewing staff
- Assurances of confidentiality and data protection
- Mentioning the survey sponsor's name
- A subject matter that is of interest to respondents
- Use of incentives.

There are some concerns that incentives can create sample bias since respondents may complete a survey solely to reap the benefit of the incentive or that respondents may view the sponsor in a more favourable light, resulting in response bias towards the sponsor's perceived desires.

However, there is little evidence to substantiate these fears. As an ONS paper<sup>24</sup> says there are, in fact, many studies which point to incentives improving data quality in terms of greater response completeness, greater accuracy, reduced item non-response and more comments to open-ended questions. Embedding randomised experiments in surveys can test for sampling bias resulting from the payment of incentives.<sup>25</sup>

For some market research methods, incentives are considered essential, for example for group discussions. For others they are strongly recommended, for example when the questionnaire is long, when the task is particularly difficult, for hall tests and depth interviews.

There are also some concerns that mentioning or not mentioning the client name may cause some bias. Mentioning the study's sponsor as part of the interview can increase response rates and improve respondent cooperation. However, for some studies, there

---

<sup>23</sup> Martin Collins, the then Research Director of Social & Community Planning Research (SCPR) says in the Consumer Market Research Handbook that although quota samples do not satisfy the conditions which define an unbiased random sample, in practice, "sample estimates are useless without some statement of their accuracy and we tend to apply the same theory. ...although there is no conclusive evidence to show that quota sampling necessarily produces estimates which are less precise than those produced by random sampling. In short, although there may be theoretical objections to applying random sample significance tests to quota samples, it is what everybody does".

<sup>24</sup> 'Incentive payments on social surveys: a literature review' Eleanor Simmons and Amanda Wilmot [www.ons.gov.uk/.../incentive-payments-on-social-surveys.pdf](http://www.ons.gov.uk/.../incentive-payments-on-social-surveys.pdf)

<sup>25</sup> See, for example, 'Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey' by van den Brakel, Jan A in Journal of Royal Statistical Society: Series A (Statistics in Society), Volume 171, Number 3, June 2008 , pp. 581-613.

may be a concern that sponsor identification could bias the results, this typically being the case with research on that sponsor's products or services. Research for government bodies and regulators do not suffer from this linkage and hence we do not believe that mentioning the CC as a survey sponsor will cause any bias.

With respect to the impact of questionnaire length on response rates, there have been a number of studies which indicate that questionnaire length in **postal** surveys has an impact on response rates<sup>26</sup>. As a general rule, long questionnaires get less response than short questionnaires. This is more likely to be an issue with postal research where non participation is much easier. With face-to-face and telephone research there is a smaller impact on non response caused by the length of the questionnaire. For online panel surveys a longer questionnaires tends to have a higher incentive which mitigates against lower response rates.

## **Marginal Customers**

A particular group of potential respondents which are often of particular interest for CC studies are marginal customers – respondents most likely to switch behaviour. The market research challenge in undertaking research with marginal customers is twofold:

- Identifying marginal customers
- Scaling their responses to the overall population.

In terms of identifying marginal customers, qualitative research approaches can help identify the nature of such respondents and help define the appropriate question or questions needed for such an identification. The research methodology then needs to be tailored to ensure sufficient such customers are sample to provide robust data.

Non marginal customers should also be sampled in the research and the market research will also need to ensure that the proportion of marginal customers to all customers is correctly identified so that weighting of the data can be undertaken.

## **Sample Size**

As a larger sample size leads to more robust data there is always a trade off in market research between cost and precision. In this respect, statistical reliability is a relative term.

The significance of data is based on a number of variables, including:

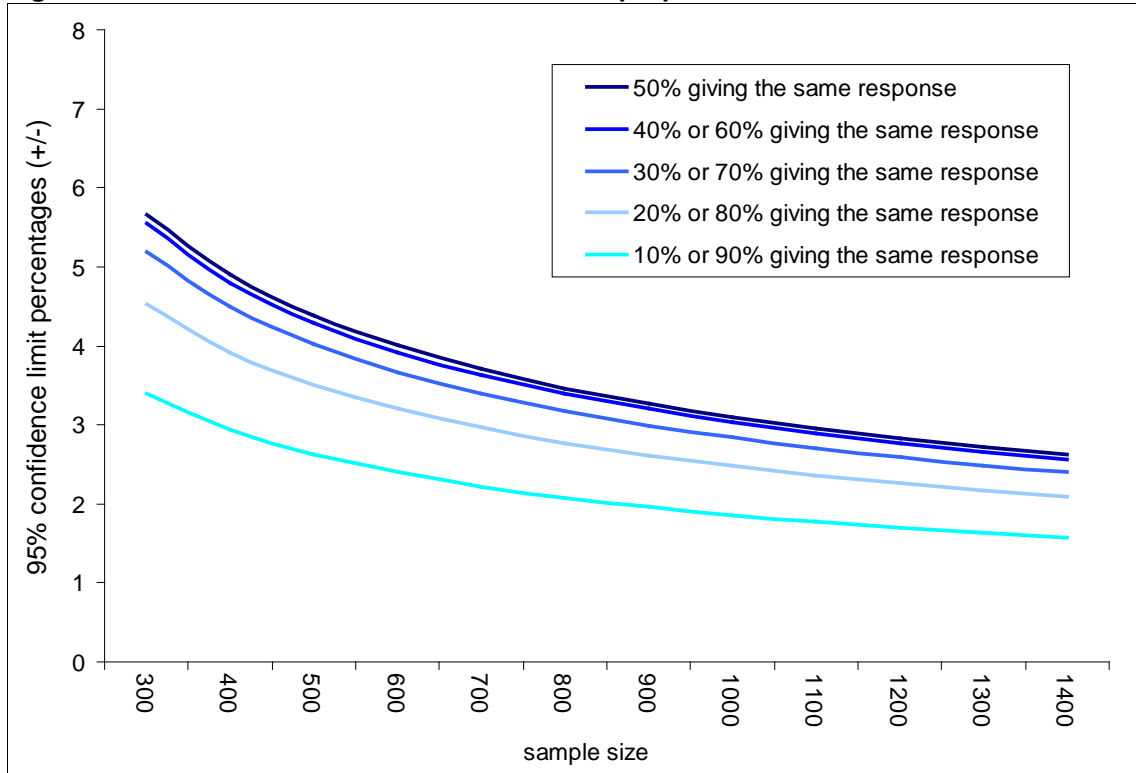
- size of population
- size of sample
- what degree of significance is required – for example 95% or 99% confidence limits.

Figure 18 provides a guide to statistical reliability for different sample sizes at a 95% confidence interval.

---

<sup>26</sup> For example: Smith Robert; Olah David ; Hansen Bruce; cumbo Dan 'The effect of questionnaire length on participant response rate: A case study in the U.S. cabinet industry'

**Figure 17: Confidence interval limits at different proportions**



For a sample of 300 if 40% agree with a statement, then we can say with 95% confidence that the true proportion in the population as a whole that agree with the statement is  $40\% \pm 5.5\%$  (that is between 34.5% and 45.5%). At 400 interviews it is  $\pm 4.8$  and at 500 it is  $\pm 4.3$ .

In market research it is common practice to use a total sample size of 400 for simple surveys, representing a reasonable balance between robustness of results and cost of fieldwork.

When specific sub groups are to be examined, the sample size for each of the sub groups needs to be large enough to provide for robust data taking account of the statistical significance when comparing data between groups.

Another important factor in determining an appropriate sample size for stated preference surveys is that the recommended size of each cell of interest is at least 75. Experience has shown that the use of smaller cell sizes are more likely to yield non significant data.

In stated preference surveys, this requirement typically means that samples are constructed on the basis of specific sub groups, each having a sample size of 75 or more. For example:

- Commuters travelling by car
- Travelling by car on employers business
- Travelling by car for leisure.

## Questionnaire Design

The questionnaire is a vehicle for the stated preference/WTP questions.

Part of the questionnaire will need to collect information that feeds into these stated preference/WTP questions, such as:

- current usage of product/service
- expenditure on product/service
- why product/service chosen
- alternatives considered, why rejected
- perceptions of product/service and alternatives.

These types of questions can both set the context for the stated preference/WTP questions and provide valuable information on the nature of respondents' decision making.

Customisation of the stated preference/WTP levels on the basis of the respondent's previous answers (such as amount paid for service/product) is an essential aspect of good design practice. Computerised approaches ease such customisation.

It is important to include questions after the stated preference/WTP questions to check that the questions are understood and have been answered appropriately. These include asking what the respondents were considering when making their choices, whether the questions were sensible and how easy the task was.

If respondents never choose one of the options or always choose the cheapest option even if the quality is much worse, then follow up questions can elicit whether there was a protest response.

The final part of the questionnaire asks for the socio-economic characteristics of the respondents. In stated preference/WTP this should include personal and/or household income as this can influence valuations.

Other characteristics such as age, sex, household structure and ethnic group are also typically collected.

## Pilot/Cognitive Testing

The pilot is an ideal that should be used whenever possible as part of the research process, particularly for stated preference approaches.

The pilot is used to test the following:

- the recruitment methodology
- the wording of the questionnaire
- interview structure
- the verbal descriptions of attributes/levels
- the show material (if used)
- the stated preference exercises (if used)
- the length of the questionnaire and whether there are any problems with respondent fatigue.

The pilot sample size for stated preference research should be at least 50.

A variant/complementary approach to piloting is the use of cognitive testing which has emerged primarily from environmental research programmes. This will typically broaden out the pilot with a series of qualitative questions intermingled within the questionnaire to probe how well the questions and scales are working.

As mentioned above, for DCE surveys pilots are important. However, due to time constraints in CC surveys, a full pilot may not be possible. In these circumstances it is recommended that the main survey is undertaken for a day or two followed by a break of a couple of days for analysis and review, before the main survey proceeds (with changes if applicable). This ‘fast track’ process is dependent on:

- use of an electronic questionnaire (ie not paper)
- questionnaire data being uploaded immediately after each shift
- close liaison with client to allow for prompt agreement of any changes.

## Quantitative Research

There are a wide range of research methods that can be used for stated preference/willingness to pay studies. These include face-to-face, telephone and self completion approaches and within each of these there is more than one method as shown below.

**Table 21: Research methods for SP/WTP studies**

Face to face	In home On street Hall tests In situ (ie on vehicle, at station, in shop, in business etc)
Telephone	Telephone Phone-post-phone
Self completion	Postal In situ (ie on vehicle) Internet

In practice, a mix of methods can be used, for example recruiting on street for a follow up telephone interview.

## Choice of survey method

The decision on which survey method should be used is often a pragmatic one, based on the nature of the required sample, the timescale and budget available.

Stated preference research benefits from the respondent being able to see the choice options so, unless the stated preference choices are very simple, this precludes telephone unless there is a means of sending the material to respondents. This method, is termed phone-email/post-phone.

Stated preference and WTP research also benefit from computerised approaches which can be applied to all the research methods shown above except postal self completion.

For on street fieldwork personal digital assistants (PDAs) can be used<sup>27</sup>, and lap top computers can be used for all other face-to-face methods. Telephone research is typically undertaken through a CATI (computer aided telephone interviews) approach.

The main advantages of computer interviewing techniques in stated preference surveys are:

- the interview can be customised to the individual respondent by carrying forward data from previous answers to be used later in the questionnaire
- routings are automatically followed, simplifying the interviewer's task and eliminating potential error
- as the need for a separate coding and data entry phase is eliminated, the reporting timescale can often be reduced and the accuracy of the data is improved
- stated preference options can be tailored to individual circumstances and the options presented can be randomised
- all data is available immediately after interview allowing for efficient questionnaire validation, sample control and field management and reporting.
- show material can be incorporated in the questionnaire.

The table below describes the advantages and disadvantages of the different approaches.

**Table 22: Advantages and disadvantages of different methods**

Method	Advantages	Disadvantages
<b>Face to face</b>		
In home	<ul style="list-style-type: none"> <li>• Allows for rigorous sampling (ie random stratified sampling)</li> <li>• Long and complex questionnaires can be administered</li> <li>• CAPI can be used</li> <li>• Allows for probing</li> <li>• Can use show material</li> </ul>	<ul style="list-style-type: none"> <li>• Most expensive approach</li> <li>• Long elapsed time to undertake as call backs required for rigorous sampling</li> <li>• Increasingly difficult to administer, particularly in inner cities and in blocks of flats</li> <li>• Incentive required</li> </ul>
On street	<ul style="list-style-type: none"> <li>• Allows for probing</li> <li>• Can use show material</li> <li>• Short timescales</li> <li>• CAPI can be used</li> <li>• Relatively low cost approach</li> </ul>	<ul style="list-style-type: none"> <li>• Sample not fully representative</li> <li>• Quotas required (but only of resident populations, not of visitors)</li> <li>• Sample over represent frequent users/visitors</li> <li>• Questionnaire has to be short (no more than 10 minutes)</li> </ul>
Hall tests	<ul style="list-style-type: none"> <li>• Long and complex questionnaires can be administered</li> <li>• CAPI can be used</li> <li>• Allows for probing</li> <li>• Can use show material</li> </ul>	<ul style="list-style-type: none"> <li>• Quite expensive approach</li> <li>• Sample not fully representative</li> <li>• Quotas required (but only of resident populations, not of visitors)</li> <li>• Sample over represent frequent users/visitors</li> <li>• Availability of locations can be limited</li> <li>• Incentive required</li> </ul>
In situ (ie on vehicle, at station, in shop, in business etc)	<ul style="list-style-type: none"> <li>• Allows for probing</li> <li>• Easy to sample appropriate market</li> <li>• Can use show material</li> <li>• Short timescales</li> <li>• CAPI can be used</li> <li>• Relatively low cost approach</li> </ul>	<ul style="list-style-type: none"> <li>• Sample over represent frequent users/visitors</li> <li>• Questionnaire has to be relatively short (no more than 10 minutes)</li> </ul>
<b>Telephone</b>		

<sup>27</sup> Although screen size means there are limitations on the complexity of the stated preference options available

Telephone	<ul style="list-style-type: none"> <li>• Long and complex questionnaires can be administered</li> <li>• Relatively low cost</li> <li>• Allows for probing</li> <li>• Short timescales</li> <li>• Easy to monitor</li> <li>• CATI can be used</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot use show material</li> <li>• Respondents may not answer sensitive questions</li> <li>• Non-telephone respondents not sampled</li> <li>• Mobile only households not easily sampled</li> </ul>
Phone-email/post-phone	<ul style="list-style-type: none"> <li>• Long and complex questionnaires can be administered</li> <li>• Can use show material</li> <li>• Allows for probing</li> <li>• Easy to monitor</li> <li>• CATI can be used</li> </ul>	<ul style="list-style-type: none"> <li>• Quite expensive approach</li> <li>• Drop out between initial contact and recontact up to 50%, so less rigorous sample and more costly</li> <li>• Respondents may not answer sensitive questions</li> <li>• Non-telephone respondents not sampled</li> <li>• Mobile only households not easily sampled</li> </ul>
<b>Self completion</b>		
Postal	<ul style="list-style-type: none"> <li>• Sampling approach can be rigorous (using PAF)</li> <li>• Low cost option</li> <li>• Lack of interviewer bias</li> <li>• Can ask sensitive questions</li> <li>• Can be completed in respondent's own time</li> <li>• Wide geographic spread easy to achieve</li> </ul>	<ul style="list-style-type: none"> <li>• Self-selection bias</li> <li>• Long elapsed time to undertake</li> <li>• No control over who completes questionnaire</li> <li>• Fixed question order</li> <li>• No probing possible</li> <li>• Restricted use of show material</li> <li>• Respondent can change earlier responses</li> </ul>
In situ	<ul style="list-style-type: none"> <li>• Low cost option</li> <li>• Lack of interviewer bias</li> <li>• Can ask sensitive questions</li> </ul>	<ul style="list-style-type: none"> <li>• Fixed question order</li> <li>• Self-selection bias</li> <li>• No probing possible</li> <li>• Restricted use of show material</li> <li>• Respondent can change earlier responses</li> </ul>
Internet	<ul style="list-style-type: none"> <li>• Low cost option</li> <li>• Lack of interviewer bias</li> <li>• Can ask sensitive questions</li> <li>• Can be completed in respondent's own time</li> <li>• Very short elapsed time</li> <li>• Wide geographic spread easy to achieve</li> <li>• Can use show material</li> <li>• Long and complex questionnaires can be administered</li> <li>• Computer applied questionnaire can be used</li> </ul>	<ul style="list-style-type: none"> <li>• Internet samples not fully representative of overall population (quotas needed)</li> <li>• Self-selection bias</li> <li>• Limited control over who completes questionnaire</li> <li>• Limited probing possible</li> <li>• Respondent can change earlier responses</li> </ul>

Face to face methods tend to have the best response rates, followed by telephone methods with self completion methods, particularly postal, having the worst response rates.

In Table 23 we show the recommended maximum questionnaire lengths. For the longer lengths incentives are typically required. If the survey subject matter is interesting then longer surveys may be undertaken and if the subject matter is boring then shorter surveys are possible.

**Table 23: Maximum recommended questionnaire lengths**

<b>Method</b>	
<b>Telephone</b>	
Telephone	<ul style="list-style-type: none"> <li>• up to 30 minutes</li> <li>• If longer, incentives required</li> </ul>
Phone-post-phone	<ul style="list-style-type: none"> <li>• up to 40 minutes (combined length)</li> <li>• Incentives required</li> </ul>
<b>Self completion</b>	
Postal	<ul style="list-style-type: none"> <li>• Up to 15 pages but marked decrease in response rates for longer questionnaires.</li> <li>• Incentives required for longer questionnaires</li> </ul>
In situ	<ul style="list-style-type: none"> <li>• Up to 10 pages</li> </ul>
Internet	<ul style="list-style-type: none"> <li>• up to 20 minutes</li> <li>• Incentives used for panel surveys</li> </ul>
<b>Face to face</b>	
In home	<ul style="list-style-type: none"> <li>• Up to an hour</li> <li>• Incentives required if over 15 minutes</li> </ul>
On street	<ul style="list-style-type: none"> <li>• Up to 10 minutes</li> </ul>
Hall tests	<ul style="list-style-type: none"> <li>• Up to 40 minutes</li> <li>• Incentives required</li> </ul>
In situ (ie on vehicle, at station, in shop, in business etc)	<ul style="list-style-type: none"> <li>• Up to 20 minutes if can be seated</li> <li>• Up to 10 minutes if standing</li> </ul>

Details on practical aspects of the methodology for each of the above methods is shown in the table below.

**Table 24: Practical aspects of different methods**

<b>Method</b>	
<b>Face to face</b>	
In home	<p>A number of sampling methods can be used for household surveys: Stratified random sample (using Post Office Address File (PAF) or Electoral Roll), random walk: 1 in n households in randomly selected streets are targeted. Individual household members can be targeted if the Electoral Roll is used. For all other methods a sampling strategy may be used to randomly sample household members, such as the 'first birthday rule'.</p> <p>Alternatively, the sample may be based on use of a relevant product or service in which case a short recruitment questionnaire would be administered to establish whether the respondent is in scope or not.</p> <p>Appointments can be made if the in scope respondent is not available at that time.</p> <p>Typically for household research up to three call backs are scheduled. Interviews are scheduled during weekdays, particularly evenings, and Saturdays so that workers can be sampled.</p> <p>Incentives are typically necessary for household interviews if they are long (say 30 minutes or over).</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
On street	<p>This method is often adopted when the research requires face-to-face interviewing with a sample of people within a specific area but where a true random sample of the residential population is not required or not appropriate. This could include, for example, the need to survey people using a specific shopping centre which would not only include local residents but also visitors to the area.</p> <p>For any on-street survey it is recommended that the interview takes no longer than ten minutes to administer. To attempt a longer interview will often result in interviews being terminated before completion.</p> <p>A short recruitment questionnaire is undertaken which probes the characteristics from the quota criteria (for example age, gender, use of product or service) to establish whether the respondent is in scope or not.</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
Hall tests	<p>The sampling approach is for potential respondents to be approached by interviewers in the vicinity of the hall. A short recruitment questionnaire is undertaken which probes the characteristics from the quota criteria (for example age, gender, use of</p>

	<p>product or service) to establish whether the respondent is in scope or not. If so, then the respondent is invited to attend an interview in a nearby hall, either then or at a time convenient to them during the hall test programme.</p> <p>Incentives are typically necessary for hall tests.</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
In situ (ie, on vehicle, at station, in shop, in business etc)	<p>This method is often adopted when the research requires face-to-face interviewing with a sample of people at a specific location.</p> <p>A short recruitment questionnaire is undertaken which probes the characteristics from the quota criteria (for example age, gender, use of product or service) to establish whether the respondent is in scope or not.</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
<b>Telephone</b>	
Telephone	<p>Sample may be provided (a user database) or purchased (typically RDD sample for a specific area).</p> <p>For resident surveys potential respondents are phoned on weekdays including evenings and at weekends. For business surveys interviews are undertaken in business hours.</p> <p>A sampling strategy may be used to randomly sample household members, such as the 'first birthday rule'.</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
Phone-post-phone	<p>Sample may be provided (a user database) or purchased (typically RDD sample for a specific area).</p> <p>For resident surveys potential respondents are phoned on weekdays including evenings and at weekends. For business surveys interviews are undertaken in business hours.</p> <p>A sampling strategy may be used to randomly sample household members, such as the 'first birthday rule'.</p> <p>After some initial questions to collect information required to prepare the stated preference options, respondents are asked to take part in follow up research and if they agree their home address/email address is taken so that a pack (including stated preference option slips and any show material) can be posted/emailed to the respondent.</p> <p>A follow up interview is then undertaken.</p> <p>The questionnaires can be administered using CAPI or paper questionnaires.</p>
<b>Self completion</b>	
Postal	<p>Sample may be provided (a user database) or drawn from the Postal Address File (PAF).</p> <p>Questionnaires, a covering letter and reply-paid envelopes are posted to respondents. One or more reminders are typically sent to non responders.</p> <p>Prize draws are often used to maximise the response rate.</p>
In situ	<p>Distribution of self-completion questionnaires to a random sample of respondents for later collection (if, for example a train survey) or for post back when questionnaires are distributed with reply-paid envelopes.</p>
Internet	<p>Sample may be provided (a user database), from an internet panel company or collected in another sampling approach (ie phone or on street survey).</p> <p>Internet respondents are typically sent an email with a web hotlink, which allows them to access the internet based questionnaire.</p> <p>Incentives are supplied by panel companies based in questionnaire length.</p> <p>Quotas are recommended to ensure the sample is representative.</p>

For all survey approaches it is good practice to record the number of people screened and found to be out of scope and the number who refuse to take part, so that an estimate of the size of the overall population can be made and to establish whether refusals are representative of the sampled population.

## APPENDIX C

### Current Practice in Competition Commission Surveys

## Current Practice in Competition Commission Surveys

This appendix comments on the market research methodology used in the case studies.

Table 25 summarises the key methodological characteristics of each of the studies including sample size, nature of respondents and the number of questions in the questionnaires.

**Table 25: Methodological aspects of individual studies**

<b>Study (Year)</b>	<b>Method</b>
<b>1. Eastman Kodak Company &amp; Colour Care Limited (2001)</b>	CAPI interviews conducted face-to-face in-home. A nationally representative sample of 2,198 adults aged 15 years or over in Great Britain. Random location sampling using national census Enumeration Districts. Data weighted to ensure that it was representative of all adults in Great Britain aged 15+. 11 questions
<b>2. First Group Plc / ScotRail Railways Limited (2003)</b>	CATI survey using RDD based upon post code sectors located around 100 train stations on relevant flows. Quotas set on age, gender and working status of adults to ensure that the sample was representative of residents in the area. 1,404 sample elicited 680 in scope respondents 50-55 questions
<b>3. National Express Group PLC/ Greater Anglia Franchise (2004)</b>	self-completion methodology amongst train passengers travelling on 'One' Anglia services between London and Ipswich, Colchester and Norwich and amongst National Express coach passengers on similar routes 547 coach and 284 train questionnaires 70-80 questions
<b>4. Acquisition of the Greater Western Passenger Rail Franchise (2005)</b>	Self completion survey administered on train with passengers travelling between specific stops. Freepost return envelope handed out. 482 FirstGroup and 529 Wessex Trains passengers. c30 questions
<b>5. Stagecoach / Scottish Citylink (2006)</b>	Face-to-face screening on coach – 3,900 contacts. 1,560 in scope, c800 volunteered for CATI and 351 did CATI interviews. On coach 30 questions, CATI c30 questions,

In general the surveys have large sample sizes and are conducted using sound methods. We make some observations on some methodological aspects of some of the case studies such as sample size, questionnaire design and sampling method.

On the First Group Plc / ScotRail Railways Limited (2003) survey over half of the 433 observations could not be used in the modelling exercise, which was ultimately based on the 196 respondents who had provided a complete set of answers for the key journey characteristics. This raises the question on why were there incomplete journey characteristics when a CATI method was used and on why there were so many outliers

On Stagecoach / Scottish Citylink (2006) the CATI interviews were conducted with just 22% of the in scope sample recruited. This raises concerns on non response bias and self selection bias.

Self completion paper questionnaires are used in two of the surveys. This is not ideal for hypothetical questions as the respondent can see all the price points. On the 'National Express Group PLC/ Greater Anglia Franchise (2004)' study it appears one self completion questionnaire was used on both routes which led to questions such as:

*“Please think about the other route that you were not travelling on”*. This could be confusing and lead to poor responses.

A very high proportion of ‘don’t knows’ (up to 39%) were given in response to some of the pricing questions in ‘Acquisition of the Greater Western Passenger Rail Franchise (2005)’. The questionnaire used per cents rather than real prices (ie *“If all of First Group’s bus fares went up by 5% (that is, 5p in every £)”*). This is likely to have exacerbated the level of don’t knows. Of course, the use of self completion questionnaires meant that the fare increases could not be calculated and printed on the questionnaire for different fares.

On all surveys there is no reporting on refusal rates so difficult to assess whether there may be an issue with non response bias.

### **Mentioning Sponsor Name**

Some of the CC research mentions the sponsor name and some doesn’t (in those cases it typically mentions the Government as the sponsor as the CC believe many people do not know who the CC is).

We would recommend that the CC is mentioned as the survey sponsor on their surveys for two reasons.

- Mentioning the sponsor name typically increases response rates, particularly if the sponsor is a government department or independent organisation as it lends authority to the survey makes respondents more likely to believe their views will be listened to. Increased response rates increase the validity of the data and reduce costs.
- If the sponsor name isn’t mentioned on CC type research, the respondent is likely to assume that the sponsor is the provider of the service or product and this, in turn, is more likely to lead to ‘political’ responses to hypothetical questions.

To allay concerns about CC not being known by the general public we would recommend that CC’s role is described, for example:

*“The Competition Commission, the competition regulator under the UK Government Department for Business, Innovation and Skills.”*

## APPENDIX D

### Current Practice in Other Competition Authorities

## Current Practice in Other Competition Authorities

A review of the current practice in three other competition authorities was undertaken as part of this study.

Telephone interviews with representatives of the US Department of Justice, the EU Directorate General for Competition and the US Federal Trade Commission were undertaken. A topic guide (see Appendix A) was used to facilitate the interviews.

### US Department of Justice

The US Department of Justice (DOJ) was represented by Norman Familant, Russ Titman and Chuck Romeo.

Market research is rarely used in the US because of the difficulty in getting it accepted as evidence in court. If a party does try to introduce a survey the other side will try to discredit it and cast doubt on every aspect of its methodology.

Although surveys are uncommon in competition cases in the US it is increasingly common for both sides to use scanner data (ie electronic database from shop scanners). If there isn't scanner data available they would use their power to get subscriber data.

When DOJ wish to answer hypothetical scenarios which demand analysis data cannot provide they use 'merger simulations'. However, they saw this approach as inherently weak as it moved outside the range of actual behaviour. Nonetheless, it has been widely used over the last 10-20 years by both sides in preparing cases. However, 'merger simulations' don't get presented in court as evidence. They wondered what would happen if they did as federal judges aren't equipped to understand a merger simulation. Critical loss analysis is used as it is more easily understood by federal judges.

Therefore, the DOJ has not done many surveys. They gave examples of two they had done:

- 1996 'Dentsply' case. The DOJ brought a monopolisation case against a false teeth distributor and charged them with entering into exclusive contracts with local regional distributors which denied other false teeth manufacturers access to local distributors and therefore the local dentists. DOJ conducted a survey which asked about dental firms' switching abilities. The survey was ignored by the judge as the response rate was too low (it was in the 30 to 40% range), no pilot was undertaken (there were concerns this would alert Dentsply and Dentsply gained access to the survey data).
- In a recent case which cannot be identified (as it *sub judice*) DOJ commissioned an extensive random telephone consumer survey of households to gain elasticities and propensity to switch between goods. They commissioned the survey on the product market as they didn't have good econometric evidence on the market to allow them to estimate substitution. The survey asked hypothetical substitution questions using a transfer pricing method. The survey first reminded respondents what they previously said they had paid before asking for their response to price changes. The DOJ felt the survey could possibly have been improved by asking, for those who

said they would switch, whether they would use the other product more or less. The survey did not get tested in court.

The DOJ thought that the latter survey gave good results and that the anti trust division would consider surveys in the future.

When the DOJ looked at what the Competition Commission used surveys for (from a paper by Graeme Reynolds and Chris Walters<sup>28</sup>), their view was that they would not have done surveys in many of those cases but instead used demand analysis data. They also thought that the response rates reported in five different surveys (10% to 46%) would not be considered acceptable in American courts.

The DOJ were very wary of using hypothetical questions. They felt respondents would give ‘political’ answers to hypothetical questions to stop the relevant company increasing prices. In this context they said that if the survey doesn’t mention who sponsors the survey it would be assumed that the company itself was asking what would happen if they raised prices, leading to ‘political’ answers.

Nonetheless, they didn’t think they could mention the DOJ as sponsors of the survey and said they understood this to be best practice.

The DOJ also has concerns on the quality of survey data on prices. They thought survey price information was “*terrible... all over the map*”. They thought that if you could get accurate supplemental price information such as distributions or data on what respondents actually paid then you could discard respondents’ prices and replace them without altering the distribution.

They referred to two sources which they used on survey methods:

- ‘Reference Guide on Survey Research’<sup>29</sup> by Shari Seidman Diamond
- ‘Sampling Methodology in Merger Analysis: The Whole Foods Survey’ by Carol K Miu<sup>30</sup>.

## **EU Directorate General for Competition**

The EU Directorate General for Competition (DG Comp) was represented by Miguel de la Mano.

DG Comp doesn’t have much experience on surveys.

The recent experience includes two airline cases (one in Ireland and in the Netherlands) and a gas retail market case in Scandinavia.

In addition they recently conducted surveys in the context of an anti trust case in the air industry which cannot be revealed as it is ongoing.

---

<sup>28</sup> [http://www.competition-commission.org.uk/our\\_role/analysis/cc\\_surveys\\_2007.pdf](http://www.competition-commission.org.uk/our_role/analysis/cc_surveys_2007.pdf)

<sup>29</sup> [http://www.fjc.gov/public/pdf.nsf/lookup/sciman04.pdf/\\$file/sciman04.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman04.pdf/$file/sciman04.pdf)

<sup>30</sup> <http://www.wsgr.com/attorneys/NEWBIOS/PDFs/walsh0509.pdf>

DG Comp relies heavily on questionnaires to gather evidence but they are not customer surveys but 'rating' questionnaires where they frequently ask hypothetical questions.

These 'rating' questionnaires are sent to a people (typically intermediate buyers, rarely customers) on a contact list provided by the merging parties. The questionnaires are long with up to 100 questions.

One customer survey they did do was on the Aer Lingus Ryan Air case as Ryan Air claimed that their passenger base was different from Aer Lingus in terms of their requirements. As it was a claim which was extremely difficult to rebut or verify they commissioned a customer survey.

From this survey they realised that there is some merit in merger and anti-trust cases in asking consumers directly on very narrow issues in situations where there is no other evidence that is available.

DG Comp has a concern on mixing factual questions with questions eliciting opinions. They thought it was difficult to discriminate between questions which give facts and those that give opinions and they felt that there could be a spill over so that customer may be influenced by the opinion questions and this might cause bias in the responses.

They thought that if they did have hypothetical questions these would preferably be in a separate questionnaire to attitude questions or for there to be the two types of questions.

Another concern with hypothetical questions was such as what would you do if the price went up by 5% or 10% was in terms of the time frame people answered it in. He believed some answered for today and others what they would have done in the past. He felt the responses would be hard to combine.

Finally, DG Comp thinks respondents may be a bit strategic in their responses to hypothetical questions and thought it was difficult to discriminate between real and strategic responses.

DG Comp are familiar with SP approaches but have not used them.

DG Comp think they will use surveys more in the future.

## **Federal Trade Commission**

The Federal Trade Commission (FTC) was represented by Mike Vita.

The FTC don't do any surveys on their own although they frequently see survey evidence in the document submissions they get in the course of a merger investigation.

The FTC thought that surveys are "*seldom right on point for the type of questions we are posing*" but they can illuminate on some product market questions and product substitute questions. Scanner data and similar RP data has a greater role than survey evidence in their cases.

Very few FTC cases go to court so he couldn't comment on whether survey data would be acceptable but he observed that it could be very idiosyncratic in terms of what a

judge chose to rely on or consider to be important evidence. As an example, he cited the Whole Foods merger case last year which the FTC lost initially in front of the trial court judge. The FTC produced a lot of evidence based on analysis of prices which carried very little weight with the trial court judge but carried a lot of weight with the Appeal Court judge where they won the case on appeal.

Mike thought that it is really hard for people to answer hypothetical questions in a credible way unless they have actually been confronted with that situation before.

FTC use academic literature in non merger cases for example when looking at a particular pricing practice that is being carried out by a firm that is arguably a dominant firm in terms of market share or where there are some distributional constraints they might use. Sometimes they hire academics if a case looks like it might go to litigation.