UNIVERSITY OF LEEDS

Updating Appraisal Values for Travel Time Savings

Phase 1 Study

Institute for Transport Studies (ITS), University of Leeds

John Bates Services

& Department of Transport, Technical University of Denmark

under the aegis of the UK Transport Research Centre

Version 1.3

20th June 2010

Report to the Department for Transport

Institute for Transport Studies

ITS

UPDATING APPRAISAL VALUES FOR TRAVEL TIME SAVINGS

PHASE 1 STUDY

Institute for Transport Studies, University of Leeds (ITS),
John Bates Services &
Department of Transport, Technical University of Denmark,
under the aegis of the UK Transport Research Centre

Authors:

Richard Batley (ITS Leeds)

Peter Mackie (ITS Leeds)

John Bates (John Bates Services)

Tony Fowkes (ITS Leeds)

Stephane Hess (ITS Leeds)

Gerard de Jong (ITS Leeds)

Mark Wardman (ITS Leeds)

Mogens Fosgerau (Technical University of Denmark)

Contents

EXECUTIVE SUMMARY

The Department for Transport commissioned a team from the Institute for Transport Studies (ITS) at the University of Leeds, in collaboration with John Bates Services and the Department of Transport at the Technical University of Denmark, to scope a possible national value of travel time and reliability study. Such a study would provide material to update WebTAG, which outlines the basis for the appraisal of transport schemes in the UK.

The case for a national study

While the team is in the position of responding to a brief, it is worth stating at the outset that we believe there is a strong case for commissioning a national study, for the following reasons.

- It is now fifteen years since the data was collected for the previous UK study (that study commenced in 1994, and reported in 1999), and it is in our opinion timely to review estimates of the value of travel time savings (VTTS). Since the previous UK study, there have been changes in travel possibilities and behaviour, policy foci, and methods of practice/analysis.

- There are open questions from the previous study, for example concerning the relationship between the value of time and journey length, group size and size of time saving, which would merit further consideration in any update study.

- We are conscious of specific policy drivers which would call for an evidence base that is broader than previous national studies. For example, the value of improving travel reliability is now seen as a much more important question with consequences for policy choices. Furthermore, in the context of the link between transport and economic performance, the values of time and reliability for the logistics sector merit review.

- At the same time as broadening the scope, we remain mindful of the importance of treating all modes, contexts and regions in a fair and comparable fashion. This would call for a co-ordinated approach to the update of VTTS, rather than the synthesis of disparate evidence bases. In our view, the latter aspiration would be best served by commissioning a single definitive national study.

- Some of the trends that we have identified, in travel possibilities and behaviour, policy, and analytical methods, could have implications for the design of an improved national VTTS study, the reporting of estimates, and the use of estimates in appraisal.

The above is based on an assumption by the team that transport cost-benefit analysis, incorporating values for the direct impacts, should remain at the heart of the appraisal process. We see travel time and reliability impacts as important proximate impacts of transport projects and also as the main channel through which transport system changes impact upon the wider economy. We believe this view is widely held among transport planning professionals, but accept that there are dissenting views.

The role of the scoping study

The focus of the brief concerns the scoping of a possible national VTTS update study. A national update study would be a substantial piece of work, and would need to inform policy and appraisal for the next decade or more. It is important, therefore, to get the study right, and this involves answering number of key questions.

- Should the core study be a repeat of the 1999 value of time study, so as to ensure comparability of results? Should it instead make use of enhanced methods of data collection and analysis, at the very least ensuring that methods are consistent with current best practice as documented by the likes of WebTAG? Or should the core study seek to combine 1999 methods with current best practice?
- While adhering to the general principle that VTTS should be comparable across modes to ensure a level playing field, is it acceptable to extrapolate values derived from car journey data (as was the case in 1999) to other modes, or should each mode be separately investigated?
- How should the study address various open questions such as the influence of group size, journey length and size of time saving on values of time savings for appraisal?
- Should the valuation of travel time reliability (VTTR) benefits be considered separately from or in conjunction with VTTS? What insights can be drawn from national studies of VTTS/VTTR in other European countries?
- Should further work be commissioned on the freight and logistics sector, and if so where would the research effort best be directed?
- Given our answers to the above, how should the main study be composed, and what are the anticipated budget, timescale and key risks?

The conclusions and recommendations of the scoping study

Arising from this study, we find as follows:

R1: There is a strong case for commissioning a revised national study updating VTTS.

R2: Any such update study should be designed around a flexible and coherent structure comprising a core study of VTTS, with provision for additional modules dealing with „areas of concern' specified by the Department.

R3: The core study should comprise two elements. The first element should be a broadly faithful repeat of the previous national study in 1999. The second element should update analysis methods to modern day standards. These two elements of the core study should be commissioned together, and from a common contractor. The study should be designed so as to be capable of providing results for the VTTS as it varies with journey duration/length and journey purpose.

R4: On balance we recommend that the update study should focus upon the car mode, and retain the assumption of the 1999 study that, for appraisal purposes, VTTS is transferable across modes.

R5: The sampling strategy for the study will need to control carefully for income so that cross-sectional variation by income can be derived and any implications for income segmentation in modelling noted. Furthermore, comparison with 1999 values will strengthen the evidence base regarding the assumed income elasticity of the value of time. From the appraisal point of view, the choice between „equity' and income segmented values is a policy choice on which the study itself will not change the arguments rehearsed in previous work.

R6: We recommend that, beyond the core study, the update study should consider three „areas of concern', namely: the valuation of travel time reliability, small time savings (STS), and the valuation of travel time savings and reliability for the freight and logistics sector.

R7: We strongly advise that VTTR be surveyed in conjunction with VTTS, broadly following the approach used in the ongoing Dutch VTTS/VTTR study.

R8: It would be prudent of the Department to additionally commission a pilot study of the „alternative' Fosgerau & Karlström approach to estimating VTTR. This would establish a contingency, should the Dutch approach prove unsuccessful.

R9: On STS, we have reservations about whether even the best designed and conducted Stated Preference (SP) study will resolve the issues of (a) whether travellers behave according to the constant unit value assumption and (b) how this should influence the Department's appraisal conventions. We have some ideas regarding alternative research approaches but are not confident that these will resolve this longstanding problem area.

R10: With regards to freight, we would advocate a strategy of making best use of the available evidence base, whilst selectively commissioning fresh analytical work where this fills knowledge gaps. To these ends, we would recommend the commissioning of an international review of methods and evidence, followed by an analysis of the effects of VTTS/VTTR changes on freight network behaviour.

R11: We recommend broadly two stages of phasing for the update study. Stage I would include review work (especially on STS and freight) and preparatory work in relation to VTTR. Stage II (involving the core VTTS survey, plus „areas of concern') would take insights from Stage I in the conduct of surveys and analysis. There remains a question as to the most appropriate staging for the freight options, which would largely be independent of the passenger VTTS/VTTR survey.

R12: At an indicative level, we recommend that the Department plans for a research programme lasting 18-24 months in duration, and makes a budget provision of £500k-£750k.

R13: In order to promote delivery on time and to budget and specification, the Department should establish clear and coherent lines of responsibility at the client end, and ensure continuity (for the life of the programme, at least) of key staff holding those responsibilities.

1 INTRODUCTION

1.1 Context of the invitation

This project arises from a Request for Proposal issued on 3rd July 2009 by the Department for Transport (DfT) to the UK Transport Research Centre (TRC). In response to this request, TRC assembled a team led by the Institute for Transport Studies (ITS) at the University of Leeds, in collaboration with John Bates Services and the Department of Transport at the Technical University of Denmark.

As we set out in more detail in Chapter 2, there have been two major Value of Travel Time Savings (VTTS) studies carried out in the UK. The first study was carried out over the period 1980-85 (MVA, ITS & TSU, 1987), while the second was conducted in 1994 (Accent & Hague Consulting Group [AHCG], 1999). The stimulus for the present project was the Department's potential interest in commissioning a survey of appraisal values for VTTS, updating the previous national study.

Mindful that any such update could be a significant undertaking, DfT commissioned the ITS-led team to conduct a scoping study (referred to as „Phase 1') of the research activities that would be involved, thereby informing the planning of any subsequent implementation („Phase 2'). More specifically, we were expected to deliver a range of options for the conduct of Phase 2, along with outline specifications for the design and resourcing of each such option. Last but not least, we were invited to issue recommendations on which options should be taken forward in Phase 2, and on the planning of Phase 2 more generally.

1.2 The case for an update to VTTS

Whilst not the focus of the Brief, we feel that it is appropriate to begin by outlining the case for a national update study, which we believe is a strong one. Our judgment is based upon the following considerations:

- It is now fifteen years since the data was collected for the previous UK study, and it is in our opinion timely to review estimates of the value of travel time savings (VTTS). We note that a number of European countries with similar appraisal philosophies have conducted national studies at roughly ten year intervals, and that some of these countries are currently in the process of completing updating studies.

- Since the previous UK study, we have witnessed substantive changes to travel possibilities and behaviour, policy foci, and methods of analysing travel behaviour. Should valuations be found to have changed in the interim, it is important to understand why this is the case, in particular isolating the influences of analytical method from changes in VTTS per se.

- There are open questions from the previous study, for example concerning the relationship between the value of time and journey length, the effect of group size and the size of time saving, which would merit further consideration in any update study. Whilst these questions continue to provoke challenges, knowledge in some areas has moved on, and it would therefore be appropriate to review the state of play, and take a view on the likelihood that an update study could deliver clearer answers.

- We are conscious of specific policy drivers which would call for an evidence base that is broader than previous national studies. For example, improving travel reliability is now seen as a much more important issue with consequences for policy choices: this requires appropriate values of travel time reliability (VTTR). Furthermore, in the context of the link between transport and economic performance, the values of time and reliability for the logistics sector merit review.

- At the same time as broadening scope, we remain mindful of the importance of treating all contexts and regions in a fair and comparable fashion. This would call for a co-ordinated approach to the update of VTTS, rather than relying on the synthesis of disparate evidence bases. In our view, this aspiration would be best served by commissioning a single definitive national study.

- Some of the trends that we have identified, in travel possibilities and behaviour, policy, and analytical methods, could have implications for the design of a national VTTS study, the format/segmentation of VTTS estimates, and the manner by which estimates are used in appraisal. Where such implications do indeed manifest, there may be good reason to collect fresh evidence, rather than draw inferences from existing evidence, especially where the latter was motivated by different interests or priorities.

The above is based on an assumption by the team that transport cost-benefit analysis, incorporating values for the direct impacts, should remain at the heart of the appraisal process. We see travel time and reliability as important impacts of transport projects and also as the main channel through which transport system changes impact upon the wider economy. We believe this view is widely held among transport planning professionals, but accept that there are dissenting views.

1.3 Setting the scene for the scoping study

A national VTTS update study would be a substantial piece of work, and it would need to inform policy and appraisal for the next decade or more. Against this background, the design of any such update study takes on considerable importance, and should address several key questions.

- Should the core study repeat the data collection and analysis methods of the 1999 value of time study, so as to ensure comparability of results? Should it instead make use of updated methods, as would constitute 2010 best practice? Or should the core study seek to combine 1999 and 2010 methods?

- The 1999 study was confined to car journeys, and, supported by the meta analysis, it was considered acceptable to extrapolate values to other modes. While adhering to the general principle that VTTS should be comparable across modes to ensure a level playing field, can we continue to rely on car journey data only, or should each mode be separately investigated?

- How should the study address various open questions such as the influence of group size, journey length and size of time saving on values of time savings for appraisal?

- Should the valuation of travel time reliability (VTTR) benefits be considered separately from or in conjunction with VTTS? What insights can be drawn from national studies of VTTS/VTTR in other European countries?
- Should further work be commissioned on the freight and logistics sector, and if so where would the research effort best be directed?
- Given our answers to the above, how should the main study be composed, what is the anticipated budget, timescale and key risks?

1.4 Options A and B

As a framework for the scoping study, the Department asked us to consider two broad options, namely:

- Option A: Review and recommend methods to update VTTS
- Option B: Option A plus attention to DfT's additional proposed „areas of concern'

The Brief articulated the „areas of concern' in terms of the matrix below, which relates 8 aspects of analytical method/functionality against 5 areas of application.

Figure 1.1: Areas of concern

Concerns	I	II	III	IV	V
Further income segmentation vs. equity values	√	√			
Group size/composition	√	√			
Trip duration/distance	√	√			√
Trip purpose	√	√			
Mode	√	√			
Small time savings	√	√		√	√
Reliability	√	√	√		√
Freight & logistics					√

where:

Category I: Represents schemes where users pay a premium for faster and more reliable journeys, e.g. road pricing and High Speed Line.

Category II: Represents of schemes for surface access to airport, where users are expected to have a higher value of time.

Category III: Represents schemes that primarily bring large benefits in improving travel time variability, e.g. Managed Motorway – Dynamic Hard Shoulder schemes and automated public transport services.

Category IV: Represents schemes that apply at a local level, e.g. smarter choices, travel plans, walking and cycling schemes.

Category V: Represents freight interventions.

In seeking to rationalise the dichotomy between Options A and B, one might recall the opening sentence of the 2003 study by ITS and John Bates Services (Mackie et al., 2003), which asserts that: *„Values of time for use in modelling and appraisal are informed by three sets of considerations – evidence, policy, and practicality'*. The authors elaborate that:

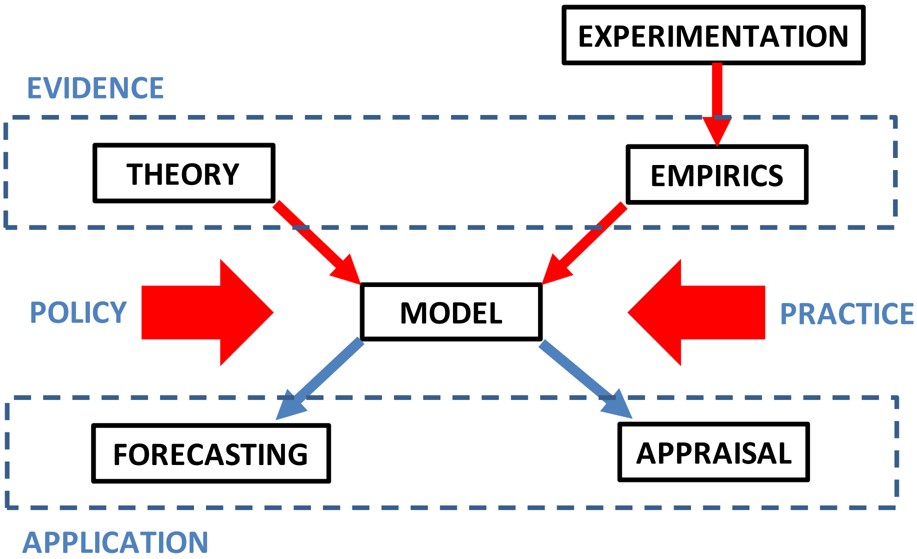
- *„The evidence may be theoretical or empirical in nature...'*
- *„In relation to policy, Governments may choose to apply VTTS in particular ways for the evaluation of public projects'.*
- *„..with respect to practicality, Government must ensure that official procedures are practical and cost-effective for the use to which they will be put'*

Figure 1.2 seeks to formalise the above assertions, representing the notion of a *„model'* as the synthesis of evidence (both *„theory'* and *„empirics'*), and *„valuation'* and *„appraisal'* as the outputs from model application.

Within this framework, one might conceive *„policy'* and *„practice'* as applying external pressures, which could influence various elements, such as the specification of the model, the scope and disaggregation of the required empirics, and the segmentation of the derived valuations.

Finally, and extending ITS/Bates' three categories of evidence, policy and practicality, our framework features *„experimentation'* as a distinct input to *„empirics'*. This reflects the influential contribution of behavioural experiments, especially Stated Preference (SP), to the current evidence base on VTTS.

Figure 1.2: Schematic framework of inputs and outputs to the analysis of VTTS



1.5 Our approach to the scoping study

The scoping study was conducted along two strands, firstly review and consultation, followed by scoping.

1.5.1 Review and consultation

This first strand was focussed around four technical reviews commissioned at the outset , namely:

- History of VTTS
- Review of methods
- Review of evidence: passenger
- Review of evidence: freight

These encompass the current state of practice (e.g. WebTAG), the current state of the art, as well as near-term emerging methods/evidence, to the extent that they could address the „areas of concern' identified by the Department. Where available and insightful, the reviews draw upon relevant experiences in other countries; for example, the Dutch and Australian experiences of valuing reliability alongside travel time savings.

An important milestone in the research project was the „stakeholder seminar', attended by invited consultants (and potential bidders for Phase 2), as well as various VTTS „users' drawn from DfT, professional associations, academia, and the transport planning community broadly. The aims of the seminar were to "ensure all experts in the group understand what we (i.e. DfT and the scoping team) are trying to achieve, and why" and to "listen to, and tap on expert knowledge and views on the challenges we are facing".

1.5.2 Scoping

Although the broad outline of the two „options' given in the Brief is clear, in practice they require considerable interpretation in order to permit more detailed consideration. In this section we set out our interpretation.

Scoping of Option A

We took Option A to constitute the „status quo' option of adhering to existing methods for estimating and applying VTTS. With reference to Figure 1.2, we would suggest that current methods arise from a number of influences, for example:

- *„Practical'* guidance in the form of WebTAG Unit 3.5.6; this is perhaps the most definitive statement of the status quo.
- Various *„theoretical'* contributions, relating to the value of time per se (e.g. Becker, 1965; de Serpa, 1971), as well as allied interests, for example reference dependent preferences (Tversky & Kahneman, 1991).
- Specific sources of *„empirical'* evidence, "primarily" AHCG (1999), but "supplemented by ancillary information" such as the National Travel Survey (1999-2001) and the New Earnings Survey (2002).
- Certain practices in *„model'* specification; for example, the AHCG study was based largely on binary logit models.
- Specific *„policy'* conventions as embodied in TAG 3.5.6, noting that the terms of reference for the scoping study explicitly excluded work travel time.

Mindful of the requirements and constraints imposed by current guidance, this report considers the extent to which value-added could be achieved by exploiting current best practice methods, as opposed to 1999 methods. With reference to Figure 1.2, we would expect such enhancements to act principally upon the modelling, empirics and experimentation elements of our schematic framework.

Having completed this assessment, we specify two alternative approaches to the implementation of Option A in Phase 2, thus:

Option A1: This would, as far as is possible and sensible, be a repeat of previous UK VTTS studies. An issue which complicates matters is that the previous UK studies have not themselves been entirely consistent, for example the MVA, ITS & TSU (1987) study considered all modes, whilst AHCG (1999) was restricted to road. In what follows, we employ a synthesised approach, combining elements of 1987 and 1999 methods. We further note that both the 1987 and 1999 studies involved a number of sub-tasks/SPs, and in

„repeating' these studies there would be the further option of how many of the sub-tasks to repeat.

Option A2: We scope out a Phase 2 study which would satisfy the same policy and practical requirements as Option A1, but exploit recent advances in methods.

For both options we develop study specifications along with additional "sub-options", prescribing the phasing of the research, the scope of each phase (including modal coverage and recommended sample sizes), and the methods to be adopted in analysis. Although the report discusses detailed features of these specifications, we should stress that Option A embodies a large number of dimensions, and could therefore spawn a vast number of potential derivatives. For this reason, the specifications described in this report should be taken as illustrative of a potential well-conceived approach to Option A, rather than the single definitive approach.

Scoping of Option B

This option is shaped by the Department's „areas of concern' (Figure 1.1). Whilst acknowledging that all of the identified „areas of concern' could potentially be relevant, we have responded to this matrix by making 3 distinctions, as follows:

1. Some of the „areas of concern' could be adequately dealt with by a carefully conceived and executed analysis, and would not especially call for dedicated study options. We would place „group/size composition', „trip duration/distance' and „trip purpose' within this category, though clearly the study design would need to recognise and plan to take on board these areas of concern.
2. We consider that two of the areas of concern, namely „further income segmentation' and „mode', are dictated more by policy convention than by analytical possibility. As regards the former, the convention is to calculate time savings for multi-modal appraisal using estimates of an „equity' value of time (i.e. not explicitly distinguishing by income distribution). As regards the latter, official VTTS are based on the car mode, and carried over to other modes. Whilst departure from these conventions would be analytically possible, this could have significant implications for analytical effort (and cost).
3. We believe that the 3 remaining „areas of concern', specifically „small time savings (STS)', „reliability', and „freight & logistics', are pertinent to current policy trends, embody significant methodological and practical challenges, and provoke a range of unanswered questions concerning VTTS policy and practice.

Following discussion with the Department, it has been agreed that „small time savings', „reliability', and „freight & logistics' should represent the foci of Option B. Whilst the composition of our report reflects those emphases, we will at various junctures rehearse the arguments for the various policy conventions (e.g. relating to „further income segmentation' and 'mode'), and scope out methods for researching all „areas of concern' more generally.

An important point running throughout Option B is that different „areas of concern' may well call for different methodological solutions, and it is quite possible that trade-offs will have to be made, between the sophistication of solutions and the degree of coverage of the areas of concern. Furthermore, it should be acknowledged that appraisal cannot take place at a higher level of sophistication than the modelling inputs which inform it. In some areas (e.g. reliability) it may be modelling capability which is the operating constraint on appraisal treatment.

Informed by the above review activities, we develop a range of study specifications, offering different approaches to Option B. These specifications offer various profiles in terms of key project management/delivery considerations, such as

- Cost
- Breadth of coverage of the „areas of concern'
- Detail/sophistication in addressing the „areas of concern'
- Practicality, including data needs and compatibility with existing methods and practice
- Risk

1.6 Layout of this report

Drawing together the above, Chapters 2-5 will cover the review strand, Chapter 6 will deal with Option A, whilst Chapters 7-9 will deal with Option B. Chapter 10 will discuss the synthesis of Options A and B in the form of a research programme, and Chapter 11 will document our conclusions and recommendations.

2 HISTORY OF UK VTTS

2.1 Introduction

The idea of empirical investigation of the trade off between travel time and money in the UK seems to have been first addressed by Beesley (1965) and Quarmby (1967), in research which built on the work by Warner (1962) in the US involving revealed preference (RP) data. It is worth noting that the analysis made use of techniques which would nowadays be considered "non-standard" (e.g. discriminant analysis), though it is unlikely that the results would have been significantly affected.

Based on stock-taking of the limited corpus of results, almost entirely using RP studies of mode choice for the journey to work, a formalisation took place among economists in the (then) Ministry of Transport. It was assessed that „the behavioural cost of non-working time be taken at 25% of the wage rate of the earner.' [MAU Note 179, 1970]: the "wage rate" was taken gross of tax. Appropriate adjustments were made to convert this to a value based on household income, so that the results could be applied to travel for non-commuting purposes, and include travel by non-earners.

For non-working time, it was agreed that a single ("equity") value should be used for cost-benefit analysis in all circumstances (except that time spent walking or waiting was given a higher value, at double the "in-vehicle" time value), whereas for working time, an early practice developed whereby separate values were prescribed for different modes and vehicles. The recommended values were published from time to time in a series of notes referred to as HEN2 (Highways Economics Note 2). The detailed calculations were not usually made explicit. Growth over time (in real terms) was assumed to be the same as for GDP (with appropriate corrections for changes in population and household size).

While both theoretical and empirical research continued, the subject reached a stage of what was described as "intellectual exhaustion", and despite criticisms from various sources, this basic methodology survived intact throughout the seventies (and was taken up, indeed, in a number of other countries).

Chief among the criticisms were a) the much higher values allocated to travel time savings in the course of work ("business"), and the practice, within the cost-benefit calculus, of valuing small time savings (literally, seconds) at the same unit rate as had been derived from empirical analysis, where the data related to much larger time savings. There was also some question about whether gains and losses should be equally valued. These issues became of more than academic interest following the publication of the ACTRA Report in 1978, which investigated the sensitivity of the CBA outcomes to a number of factors, including the differences between working and non-working time. It also commented at some length on the issue of small time savings, though on balance it supported the *status quo*.

2.2 The first UK value of time study

In 1980, at least partly in response to the ACTRA report, the (then) Department of Transport commissioned a new study, though there were substantial delays in the procurement process for the various stages and it only reported in 1987 (MVA, ITS & TSU, 1987)). The study had a wide remit, and was partly supported by British Rail, so that values for all modes were in the remit. Although in the course of the study's development it was decided to exclude working time values from its remit, a separate ESRC-funded study investigated this in parallel (Fowkes et al., 1986).

One of the key requirements was to widen the empirical base to include other leisure purposes. However, it was quickly realised that the conditions for obtaining values from **RP** data were prohibitive. Hence the study experimented with forms of hypothetical data, including Transfer Price (Contingent Valuation) which had previously been used by Lee & Dalvi (1969), and SP.

Based on the conclusions of these experiments, this was the first VTTS study to rely substantially on SP data, though some RP data was included. The choices offered related both to mode choice and route choice. It may be noted that, despite the theoretical developments carried out in the course of the study, the practice of SP design, presentation and analysis was relatively rudimentary by current standards. Nevertheless, the study was able to conclude: "The experience of this approach [SP] built up during the study has been generally favourable: the results have been comparable with those obtained by the more traditional [ie, RP] approach, and the variations in values of time obtained from the analysis have accorded well with expectations."

The study also made substantial progress in the theoretical account of time valuation, building on previous academic work by (among others) de Serpa (1971) and Bruzelius (1979), while the analysis made extensive use of the discrete choice paradigm developed in the 1970s.

The Department generally accepted the findings, though there was controversy about whether the values obtained from drivers should be viewed as values for the whole vehicle (as recommended in the Study Report) or the driver and then extended to the passengers (the Department's interpretation). They issued a new version of HEN2 in May 1987. Subsequent versions[1] of the note did not make significant changes in the Value of Time methodology.

The headline outcome of this work was that the Department's philosophy of evaluation, including the standard value of non-working time, was retained intact, but the standard value itself was increased by 58 per cent to 43 per cent of the average hourly earnings of full time adult employees, which was equivalent to 40 per cent of the mileage weighted hourly earnings of commuters.

For travel on employers' business, the ESRC study (Fowkes, Marks & Nash, 1986) had attempted to populate (and, indeed, reformulate) the so-called „Hensher' equation (Hensher, 1977) but a range of results were found (depending on uncertain inputs) which overlapped those from the traditional „cost saving' approach, which was therefore retained, with recommended values for categories such as bus and coach drivers, commercial vehicle drivers and car drivers on employers' business. Thus the value is assumed to be proportional to the wage rate (or income per hour). For each chosen category, the mileage

---

[1] Following the version issued in 1997, the series was replaced by the "Transport Economic Note" [TEN] first issued in March 2001.

weighted average annual income can be calculated, using NTS data for journeys subdivided by income and journey length, as well as other parameters of interest (e.g. mode). An allowance for overheads of 24.1% is then added, to give the total annual cost to the employer. The value of working time is then derived by dividing by the total number of hours worked, obtained from the New Earnings Survey[2].

2.3 The second UK value of time study

In 1994 a new study, specifically for highway values, was carried out by Accent Marketing and Hague Consulting Group [AHCG], though it did not officially report till 1999. This study adopted a general methodology similar to that underlying the first Dutch study (HCG, 1990): it was based entirely on SP data. Although three different forms of SP experiment were used, the majority of the recommendations were based on "Experiment 1", which involved a direct route choice trade-off between time and cost changes from a reported base position. Most of the data related to car travel, but lorry and coach journeys were also investigated. The study attempted to apply a variant of the Hensher methodology for Work travel, combining both employer- and employee-based valuations.

The analysis was carried out at considerable detail, and identified a wide range of "co-variates" influencing the values of time, though for practical purposes, averages were calculated for three purposes (Business, Commuting, Other leisure). Investigations were made into the "size and sign" of time savings, and it was concluded that VTTS was **not** independent of the direction and magnitude of the time saving (though, once again, averages were provided for practical purposes).

The Department had some difficulties in responding to the conclusions of the study, and in 2001 the University of Leeds Institute for Transport Studies, together with John Bates Services, were commissioned to help resolve some of the outstanding issues. This involved a substantial re-analysis of the AHCG data. The ITS/Bates study reported in 2003. With respect to the calculation of VTTS, it made a series of recommendations (references in square brackets refer to the relevant sections in the Final Report). Note that, although some analysis of co-variates was undertaken, given the perceived difficulties in the AHCG report of both presenting and implementing these results, the main conclusions were based on models which generally omitted such effects.

In the first place [§5.4] the recommended model of non-working VTTS for the purposes of demonstrating the variation in values was:

$$\text{VTTS} = [\beta_T/\beta_C].\left(\frac{Inc}{Inc_0}\right)^{\eta_{Inc}}.\left(\frac{C}{C_0}\right)^{\eta_C}, \qquad (2.1)$$

where Inc is household income in £'000 pa and C is journey cost in pence (both in 1994 prices), and Inc_0 and C_0 take the fixed values 35 and 100 respectively. The estimated values

[2] From October 2004 the New Earnings Survey (NES) was replaced by the Annual Survey of Hours and Earnings (ASHE).

for the parameters β_τ, β_c, η_{Inc}, η_c are given in the table below (with t-ratios in brackets), separately for the Commuting and "Other" purposes:

Table 2.1: Estimated parameters for time, cost, income elasticity and distance elasticity

	Commute	Other
$\Delta\tau$ β_τ	-0.10098 (-15.07)	-0.08292 (-19.33)
Δc β_c	-0.02473 (-14.84)	-0.02227 (-18.51)
Income Elasticity η_{Inc}	0.35878 (7.58)	0.15681 (5.49)
Distance Elasticity η_C	0.42130 (9.08)	0.31473 (11.86)

In principle this formula allows the VTTS to be calculated for any combination of traveller income and journey **cost**. The actual distance travelled was not available in the AHCG data. Nonetheless, it was clear that the sample distribution was not representative [§5.2]. It was therefore considered essential to construct a "bridge" between the AHCG reported costs and journey distance, both to correct for representativity and to find a way of applying the preferred model in relation to data sets which do not contain the relevant information on cost.

For this purpose, a relationship was derived [§5.3] which suggested that the cost per mile had the average value of 13.2 pence (1994 prices), and that for most of the data a linear assumption between distance and (reported) cost was justified. On this basis, C in the VTTS formula can be interpreted as distance (in miles) and the value of C_0 can be converted to D_0 = 100/13.2 = 7.58 miles (12.2 Km).

To correct for representativity, the NTS was chosen, since it classifies all journeys by journey distance. DfT was able to provide tabulations from NTS for the six years 1995/2000 giving information on the distribution of all the variables of interest (that is, income, distance, purpose and mode). This allowed the chosen model to be applied directly to the NTS data for 252 "cells", based on household income in 21 bands, and journey distance in 12 bands.

Based on the "meta-analysis" [§6.8] which built on an approach developed by Wardman (2001), it was recommended that while for working time, growth should be in line with income, for non-working time an elasticity of 0.8 should be used. It will be noted that this is a higher elasticity than was revealed in the cross-sectional analysis.

It will be seen from Table 2.1 that the distance elasticities are larger (and more significant) than the income elasticities. In addition, it turns out that the range of journey lengths in the NTS is more skewed than the range of (household) incomes. It is also important to point out that, while the NTS data reveals some correlation between income and journey length, this does not have a large effect on the average value of time.

In carrying out the calculation of representative values, three significant issues were identified [§7.3]:

- model stability - treatment of income growth
- modal variation

- weighting - (trip vs distance)

It was assumed that, apart from corrections for the price level in which income is measured, the models for the VTTS relationships could be taken as temporally stable, so that they could be applied to data for years other than the one in which the underlying data was collected (end-1994). The NTS data related, on average, to a later year than the estimated VTTS relationship, and it was argued at the time that no correction needed to be made for the fact that some income growth would have occurred[3]. As will be seen, subsequent analysis indicated that this was not, in fact, appropriate.

On modal variation, while there was some empirical evidence for modal variation in VTTS which generally accorded with intuition (in the sense of reflecting presumed effects of comfort, cleanliness, information and other modal characteristics), it was not particularly strong. On this basis it was assumed that the estimated car driver VTTS models could be applied to all (mechanised) modes (in respect of in-vehicle time).

On weighting, it was judged that a distance-weighting was preferable to a trip-weighting. This coincided with the Department's conventional approach in the past to **working** values of time, discussed earlier. However, in the case of non-working time, it was unclear what previous practice had been: in any case, while the identified variation might cover a large range of variables, in the 1980s UK study it was relatively limited. It was noted that the fact that, in broad terms, journey length tended to increase values of time might suggest that by using a distance-weighting, there is an element of double-counting: certainly a distance-weighted average will be higher than a trip-weighted average. Nonetheless, it was argued that a principal aim should be to allow time savings to be aggregated in a way which coincides with the total "value" reflected in individual values of time.

In other words, if N_{yd} is the number of trips in the NTS data that fall into the category of income group y and distance group d, and V_{yd} is the corresponding VTTS according to the formula given, then to calculate the average for the combinations $\{y \in Y; d \in \Delta\}$, where Y and Δ are grouped categories of income and distance respectively, we use the formula:

---

[3] As a result, the implied formula could be written as:

$$K\,[\beta_T/\beta_C].\,\left(\frac{Inc'}{Inc'_0}\right)^{\eta_{Inc}}.\left(\frac{D}{D_0}\right)^{\eta_C}$$

where *Inc'* represents the NTS household income in £'000 pa, D is NTS distance (in miles), D_0 is set to 7.58 miles, Inc'_0 is set to the value 35*K, and K is correction for inflation between the last quarter of 1994 and the average date to which the income data applies. Since the average RPI value for the last quarter of 1994 is 145.5, while the value for December 1997 (end of the first three years in NTS) is 160.0, the correction for inflation (K) was taken as 1.100.

$$\bar{V}(Y, \Delta) = \frac{\sum_{y \in Y} \sum_{d \in \Delta} V_{yd}.N_{yd}.D_d}{\sum_{y \in Y} \sum_{d \in \Delta} N_{yd}.D_d}$$

where D_d is the average distance for distance band d.

Since the relationship of VTTS with both income and distance is quite non-linear, it is **not** appropriate simply to input average income and distance into the VTTS formula.

It was also considered [§8.1] that for **modelling** purposes there was some case for considering the possible use of values of time varying at least with income and distance. On general grounds of practicality, this was restricted to three groupings for each variable, leading to 9 cells in all. Boundaries were drawn based on the following considerations:

a) the existing boundaries of the NTS tables should be respected;
b) approximately 1/3 of the sample should fall into each variable grouping;
c) the grouping should be the same for both purposes.

For various reasons, this level of detail was not included in the revised TEN Note which was issued in 2004. In other respects, however, the recommendations were accepted. Overall average VTTS for non-working time made use of the formula and weighting process described above, and for changes over time, an elasticity of 0. 8 would be applied to income growth (GDP per head) rates. Note that the effect of this will be for the value of time to reduce over time as a proportion of personal income.

Despite the considerable efforts made by AHCG to implement values of **working** time using the Hensher formula, ITS/Bates concluded that "the Hensher formula approach is data hungry, and that none of the various parameters r, p and q are at all easy to estimate with confidence. AHCG have made a fair attempt, but their basis for imputing the MP values is weak. Therefore we cannot recommend adoption of the approach taken and values derived in the AHCG report." [§3]

ITS/Bates argued that, in fact, the (traditional) assumption of valuing working time savings at the marginal cost to the employer (wage rate plus an uplift for overheads) is reasonable for many categories of workers, such as commercial drivers and travelling salesmen. It is least supportable in the case of business ("briefcase") travellers, where the implications of the way in which time savings may be converted to leisure are most apparent. In this case, with some nervousness, ITS/Bates concluded that "given these uncertainties, and given that alternative assumptions give results either side of the Department's current value, we see no strong case for abandoning the cost saving approach for valuing savings in travel time for briefcase travellers". As a result, the overall methodology for working time values (described earlier) was unchanged.

As discussed, the average non-working values given in the ITS report were interpreted as being end-1997 values in end-1997 prices. For the TEN, these values were converted to 2002 values and prices by uprating in proportion to changes in values of time growth and changes in prices (using the Retail Price Index) between end-1997 and mid-2002.

Subsequently, the role of the TEN was taken over by WebTAG Unit 3.5.6. At the time of writing, the values remain unchanged, at 2002 values, however.

In line with the recommendations from the ITS/Bates report, no adjustment was made for the growth of real GDP/head for the three years 1994-1997, on the grounds that a) it was a small effect, and b) that it was expected that at least some of the growth would have been implicitly taken into account by changes in the distribution of later NTS data across income and distance bands.

2.4 Subsequent work

With the passing of time, more recent NTS data has become available which can be used to reweight the VTTS for the years 2002-2005.

It is therefore now possible to see how much a change in the weights contributes to change in the average cross-sectional VTTS. Rather surprisingly, after adjusting for inflation, there is virtually no change in the average VTTS as a result of the new NTS data. This turns out to be because the distribution of trips by distance and (real) income is virtually unchanged between the two NTS data sets.

As a result of this, it may be concluded that the 1997 non-working values reported in the ITS/Bates report **should** have made allowance for the growth in GDP per head between 1994 and 1997, and that the VTTS formula should be construed as providing **1994** values (in whatever prices are represented by "K"). This means that any averages obtained by this formula need to be further adjusted for the growth in GDP/head from 1994, with the recommended elasticity of 0.8.

Hence, the formula should be re-written as:

$$G^{0.8}.\ K' \left[\beta_\tau/\beta_C\right].\ \left(\frac{Inc'}{Inc'_0}\right)^{\eta_{Inc}} . \left(\frac{D}{D_0}\right)^{\eta_C}$$

where, as before, *Inc′* represents the NTS household income in £'000 pa, D is the trip distance, D_0 has the value 7.58 miles or 12.2 Km, according to the units chosen for D, and Inc'_0 is set to the value 35*K, K being the correction for inflation between the last quarter of 1994 and the average date to which the income data applies.

K′, by contrast, is merely an inflation factor between the last quarter of 1994 and whatever year's prices are chosen to represent the values of time in.

G represents the real growth factor in GDP per head between 1994 and the year to which the VTTS is to relate: thus if GDP has grown by 20%, G has the value 1.20.

This formula is intended to be applied to disaggregations of trips by quite detailed income and distance categories (as in the NTS data). To obtain averages for more aggregate categories, the distance-weighted calculation is required, as described above.

This change in methodology means that the WebTAG 3.5.6 values of time are not quite correct (see also WebTAG 3.12.1, where the revised formula is applied for the purposes of income segmentation). However, the difference is small, and for the purposes of continuity, no decision has been taken by DfT to revise the values at this stage.

2.5 Issues arising

There are four key issues which have arisen consistently in all VTTS studies, and will not go away. Some of these have already been alluded to in the previous discussion. They are:

- small time savings
- differential valuations for gains and losses
- valuation of business traveller time (although we remind the reader that this is excluded from the remit of the current scoping study)
- value of freight time

In addition, there is a further issue which has attained more prominence recently, though it, also, is of long standing:

- value of travel time reliability

In this section, we will briefly summarise the current position on each of these items. Finally, it is worth making some general remarks about modal variation.

2.5.1 Small time savings

There is a basic conflict here between the demands of the appraisal system, which requires that the value of a scheme or package should not be dependent on arbitrary disaggregation, and the common sense view that a change in travel time of a small amount (well within the threshold of uncertainty or possibly even perception) should not be allocated value. This is exacerbated by the fact that in many cases the make-up of total time savings attributed to a scheme does indeed consist of a large number of very small savings.

The AHCG study produced evidence that, in terms of SP choices, the unit value of time was smaller when small changes in time were offered. Of course, there is the recurrent question of what we mean by „small', and in the case of AHCG we are referring to time savings of 5 minutes or lower. More generally, the ITS/Bates study counselled against taking this result at face value [§4.2]. They noted that: "In the first place, the results are inconsistent with the theoretical expectations on the shape of the indifference curve, at least when allowance is made for adjustments beyond the immediate short term……. On balance, we feel that the lower values for the small time savings arise because of the artificial nature of SP exercises, and the large imaginative leap the respondent is required to make to answer the question in a long-term rather than an immediate term manner". Nonetheless, there is an element of discomfort in the argument. A recent contribution in this area is the work of De Borger & Fosgerau (2008), who offer a theoretical explanation for the STS phenomenon.

Although it remains unclear whether any progress can be made, either empirically or intellectually, it is accepted that the matter needs to be addressed.

2.5.2 Differential valuations for gains and losses

Although this is an equally well-discussed issue, it is perhaps more easily dealt with. There is considerable reported evidence that improvements may be less valued than equivalent losses of the same amount (though it may be noted that the AHCG results were rejected in this respect by ITS/Bates because of the failure to account for the *status quo* bias in the SP responses). This leads to the unfortunate conclusion that giving someone an improvement and then removing it (or *vice versa*) will leave them worse off than at the start. While this may have a small amount of psychological relevance, it is not compatible with conventional economic theory. For this reason, the empirical evidence has generally been rationalised as essentially a short term response, the (slightly) longer term being required to make the adjustments consequent on the change to the (time or money) budget.

From this viewpoint, the issue is really how to interpret the SP responses, **in the light** of the likely short-term bias.

2.5.3 Valuation of business traveller time

As is clear from the earlier part of this chapter, accepted practice for the valuation of business traveller time is on a different basis to that of other travel time, in that it is not based on the individual's willingness-to-pay (WTP), but on the presumed WTP of the employer (assumed, moreover, to be equivalent to the marginal productivity as measured by the wage rate). It has always been accepted that this requires a number of conditions to apply to ensure validity.

A "rival" theory, the "Hensher" formula, is more or less unassailable, though there can be discussion regarding the terms included and their exact definition. It allows for the possibility that a travel time reduction could lead to a reduction in time spent productively working while travelling, or in the total time spent working and travelling. Our current understanding (the same as in ITS/Bates (2003)) follows the developments introduced by Fowkes, Marks & Nash (1986). These were not followed by AHCG, with consequent errors in their study, and in some other studies that followed their method. All attempts have found it impossible to reliably populate the Hensher equation. Fowkes et al. used ranges, which overlapped the value from the "cost plus" approach. For these two reasons, ITS/Bates did not feel it sensible to recommend a departure from the conventional approach. However, since then, more empirical work has been carried out, particularly relating to "briefcase" travel by long distance rail.

It would therefore be worth reviewing this topic, and considering whether there is scope for moving towards at least a modified version of the Hensher formula, if appropriate measurements can be conducted. The review would also provide the opportunity to address some of the misconceptions prevalent with this topic (for example, the idea that the phrase "in working time" has any relation to official hours of work, or that train time should have a **higher** value because it offers greater scope for productive use!).

2.5.4 Value of freight time

The current UK convention is that freight travel is treated essentially in terms of the driver (and passengers, where relevant). Thus, in line with the general practice regarding "business" travel time, changes in freight travel time are valued at the driver's wage rate (along with the standard impact on vehicle operating costs).

From this point of view, the same issues arise as in relation to business travel, except that the marginal productivity approach is much less controversial in the case of professional drivers. However, there is also the question as to whether the cargo has an implicit value of time: it has been proposed, and implemented in some countries (for example, Sweden), that it is appropriate to apply a capital-related charge, given the forced unproductivity of the cargo during transport. In addition, there are issues relating more to punctuality, in connection with perishable commodities, just-in-time production processes etc.

The concomitant question is whether any or all of these aspects can be deduced from economic principles (given appropriate data, of course), or need to be elicited in the form of WTP surveys (for example, using SP techniques). Although there is a growing corpus of SP and other work in this area, there remains some dispute about its relevance and reliability.

2.5.5 Value of travel time reliability

The 1980s study proposed reliability as an important area of investigation, but as part of the resource budgeting, a full treatment was excluded from the research programme. "Hence such indicative results as we have obtained have emerged from our design of surveys rather than from any major analytical effort." [p175]. It was argued that the evidence suggested a "sliding scale" whereby values of time could increase under conditions of unreliability (which, in the case of highway travel, was particularly associated with congestion).

Since then, much more work has been done in this area, including recent review work for the Department (ITS, Imperial College & John Bates, 2008). Indeed, the Department has subsequently released guidance on VTTR (WebTAG 3.5.7) alongside existing guidance on VTTS (WebTAG 3.5.6). Unlike its VTTS counterpart, standard values of reliability detailed in WebTAG 3.5.7 do not arise from a single definitive national study, but are instead drawn from a range of sources. Key sources include Black & Towriss (1993) for road, Bates et al. (1990) for rail, and Atkins (1997) for bus, although the „headline' results of these studies have been broadly supported by subsequent work.

Given the vintage of the aforementioned sources, there is however a good case for wishing to review (and if necessary revise) VTTR guidance on the basis of updated evidence, especially where this evidence arises from a national study combining VTTS and VTTR. There is an increasing tendency to carry out studies which **jointly** attempt to derive values of time and reliability (ongoing work in Holland, Norway). A case can be made for introducing this approach in the UK, since this would permit economies in the costs of survey work, and would serve to expose any duplication/confounding between estimates of VTTS and VTTR.

A particular problem has been the presentation of different levels of reliability in the SP context. An alternative approach is to accept the general implications of *scheduling theory* (building on the work by Vickery (1969) and Small (1982)) under conditions of uncertainty, and restrict the SP choices to those relating to early and late arrival, under conditions of **certainty**.

In any case, there is scope, within this context, for a re-evaluation of the overall theoretical model for time allocation. A particular issue which the 1980s study was not able to address was the "constrained transferability of time". Given the recent work in this area (a particular interesting generalisation of the scheduling model is given in Tseng & Verhoef (2008), based on earlier work by Vickrey (1973)) this could be a fruitful avenue, for the valuation of both time and reliability.

2.5.6 Modal variation

Section 8.3 of ITS/Bates (2003) provides a general discussion, as well as a commentary on some of the empirical findings presented in earlier sections, in particular §7.3. We begin with a summary of the main points.

Average values of time by mode may vary for reasons which are partly connected with the socio-economic characteristics of the users and the kind of trips that they make, as well as for reasons inherent to the mode, which may be broadly described as “comfort (or quality) effects”. It is important to separate these two contributions, but limited evidence was available from observations contributing to the 2003 “meta-analysis”.

These observations related to cases where valuations for time savings on one mode had been estimated from (current) users of another mode. This allowed a cross-classification of the VTTS from the same sample with respect to different modes. When no account was taken of this effect (i.e. when the VTTS was confined to current users of the mode), the general pattern emerged that $VTTS_{bus} < VTTS_{car} < VTTS_{rail}$. But when the analysis controlled for the sample of users, exactly the opposite effect was found: $VTTS_{rail} < VTTS_{car} < VTTS_{bus.}$ This latter result was interpreted as variations in VTTS by mode caused by variations in the valuation of comfort, cleanliness, information and other modal characteristics (as distinct from variations in the characteristics of the users).

In spite of this result, the statistical evidence for the variation due to “comfort” was not considered sufficiently robust to justify a differentiation of VTTS by mode on this account. (As we shall see subsequently in Chapter 4, an updating of our meta-analysis has not served to strengthen the case for variation due to comfort. Indeed, quite the opposite; the evidential picture is arguably weaker as a result of a bolstering of our evidence base). Working therefore with a single model for all modes, the 2003 report demonstrated that the observed variation in mode use by income and journey length would produce variation in **average** values for mode used in line with the results from the meta-analysis, giving some assurance that the commonly reported results were essentially due to variations in the characteristics of mode users and the journeys that they make.

Subsequently §8.3 went on to address the question "is there a case, in the light of the theoretical and empirical evidence, for modally differentiated values of non-working time?". It concluded that the use of values taken direct from existing *users* of each mode would lead, both in demand modelling and in appraisal terms, to inconsistencies, essentially because “switchers” between modes will be treated anomalously. The logical way forward was to allow explicitly for the different income composition of the travelling population, with appropriately differentiated values of time.

By contrast, no such problems would arise from allowing **the same individual** to have different modal values of time, providing that these were essentially reflecting variations in comfort and convenience. However, as there would be some increase in complexity, it should only be contemplated if the evidence for such differentiation can be made sufficiently strong. As noted, the ITS/Bates (2003) study concluded that this was (at that time) not the case. However, this is essentially an empirical issue, though if it were to be investigated, it would be important to show that the selected methodology would be able to reveal a difference, if it existed.

Note that there would be little value in conducting entirely separate VTTS studies on samples of the users of different modes, since this would merely reproduce the modal variation due to the sample make-up. As a minimum for identifying the “comfort” contribution, it would be necessary to control for all identifiable effects contributing to variation in VTTS, and it would be much more desirable to present users of current modes with specific comparisons related to other modes (as was the basis of relevant evidence from the meta-analysis).

The conclusion we have arrived at is that before commissioning work on measuring the modal attribute values for use in appraisal, the Department needs to consider a number of related issues. First and foremost, of course, is the question as to whether the results are likely to be sufficiently robust and generic for practical application. However, even if they are, there is still the issue as to whether they would add value to the results already available from the meta-analysis of many studies.

There are further policy questions, particularly those relating to appraisal guidance (WebTAG). Would the use of mode-specific values be optional or mandatory? Could the central values be overridden by local project-specific values? There would also be a need to resolve potential incompatibilities with PDFH for rail applications.

Finally, it would be useful to consider some of the modelling issues, given that, although value of time guidance relates essentially to appraisal, there is a strong tendency for models to ensure compatibility by using the same values for demand forecasting. Thus, for example, one might wish to be sure that no anomalies would arise from using mode-specific values when considering busways versus bus routes on the highway.

2.6 European perspectives

2.6.1 National VTTS studies

Finally, it is useful to set the UK experience against similar experiences from other European countries. As a basis for this discussion, we consider four specific countries (The Netherlands, Sweden, Norway and Denmark) which, like the UK, have established a tradition in the conduct of national VTTS studies. The Netherlands has been a leading contributor to the evidence base, commissioning passenger VTTS studies in 1988-1990, 1997-1998 and 2007-2010, and freight VTTS studies in 1990-1992, 2003-2004 and 2007-2010. Sweden is also notable in having undertaken repeat studies for both passenger (1996, 2007-2009) and freight (1992, 1999 and 2009-2010). Similarly, Norway has undertaken a repeat passenger study (1997 and 2008-2009), and has recently undertaken its first national freight study (2008-2009). Last but not least, Denmark's contribution has been a single passenger VTTS study (2006-2007).

2.6.2 Method and evidence

In what follows, we summarise the key features of method and evidence associated with the national VTTS studies of these four countries; for further detail the reader is referred to Annexes A-D of the present report.

Dealing first with evidence, we refer to Table 2.2, which summarises the segmentations/covariates of interest within the national VTTS models of the four countries. With regards to passenger, it can be seen that mode and purpose are standard segmentations in all four countries, whilst 3 of the 4 countries also segment by distance (the exception being The Netherlands which instead segments by speed). Size and/or sign of time savings are considered by all except Norway. On freight, the picture is much less consistent, with only The Netherlands and Sweden having reported mode-specific valuations, and the Norwegian analysis still ongoing.

Table 2.2: European evidence for VTTS

	NL	S	N	DK
Passenger VTTS varies by	Mode, purpose, income, speed, sign and size of savings	Mode, purpose, income, time, distance, size of savings	Mode, purpose and distance. New model estimation is ongoing	Purpose, income, congestion, time, distance, sign and size of savings
Freight VTTS varies by	Mode, road: raw/final goods, value of goods, containers	Mode	Model estimation is ongoing	N/A

Turning to methods, we refer first to Table 2.3, which summarises the „standard' approaches adopted in the four countries for VTTS. As regards working time (which, as noted, falls beyond the remit of the current study), there is a dichotomy of approach, with The Netherlands and Norway adopting the Hensher method, and Sweden and Denmark the „cost savings' approach. On non-work there is more uniformity, with all countries employing measures of willingness-to-pay for in-vehicle time (albeit with some differences in segmentations). Finally, there is no clear consensus on VTTS for freight, with each country adopting a different approach.

Table 2.3: European methods for VTTS

	NL	S	N	DK
Work VTTS	Hensher method, by mode (and income); employee part from SP	Cost savings, by mode	Hensher method by mode and distance	Cost savings
Non-work VTTS	WTP to save IVT, by mode, purpose (income)	WTP to save IVT, by distance	WTP to save IVT by mode, purpose and distance	WTP to save IVT
Freight VTTS	WTP, by mode (and type of goods)	Interest cost+[4], by mode and commodity	Value of alt. use (+cargo component), by mode	Cost savings +cargo component, by mode

[4] This refers to components of VTTS other than the interest cost on the capital invested in the goods that are being transported, but that are also related to the cargo itself, not to the provision of transport services. This would include deterioration of the goods during transport, disruption of production processes, and running out of stock at the receiving end of the shipment.

Proceeding finally to Table 2.4, which summarises the „standard' approaches for VTTR, there is a further dichotomy between applications of the „reliability ratio' (The Netherlands for all passenger and freight; Sweden for car only) and the notion of delay time (all other modes/countries). We will pursue these methodological possibilities in more detail in Chapter 9, which scopes out methods for the valuation of reliability.

Table 2.4: European methods for VTTR

	NL	S	N	DK
Passengers	Reliability ratio by mode, from expert workshop (or 20% mark-up[5])	PT: delay time[6], by purpose and distance Car: Reliability ratio by purpose	Delay time	Rail: delay time
Freight	Reliability ratio by mode, deduced from SP study	Delay time	Delay time	Rail: delay time

2.7 Conclusions

This chapter has outlined the historical context of VTTS in the UK, and made an important clarification regarding the application of recommended values, as prescribed by WebTAG 3.5.6. As regards the „areas of concern', it was acknowledged that a number of issues, specifically STS, VTTR and freight VTTS/VTTR, are pertinent, but difficult to analyse. The issue of differential valuations for gains and losses was judged to be less critical. Last but not least, the valuation of business time was deemed to be worthy of further analysis, but falls outside the terms of reference for the present study, and will not therefore be discussed further.

Contrasting UK practice against practices adopted in 4 other European countries (The Netherlands, Sweden, Norway and Denmark), it is notable that all of these countries employ a „standard' approach of valuing non-work VTTS using WTP methods, albeit with some differences in the segmentations used. There is less unanimity on VTTR for passengers, although all countries use either or some combination of the reliability ratio and/or lateness as the relevant metric. Freight VTTS shows considerable variety in method, whilst freight VTTR shows more consistency, with most countries focusing upon delay as the relevant metric.

Having reviewed the historical origins of VTTS, and briefly charted the evolution of methods and evidence from the 1960s up until the present day, the following three chapters will

[5] The 20% mark-up to VTTS arises from a standard reliability ratio (VTTR/VTTS) of 0.8.

[6] Delay time is defined as journey time in excess of free flow journey time or scheduled journey time, depending on context.

review methodological and empirical evidence in greater detail. Chapter 3 will deal with methods, whilst Chapters 4 and 5 will deal with evidence (the former focussing on passenger, and the latter on freight).

3 REVIEW OF METHODS

3.1 Introduction

This chapter discusses survey and estimation methodologies in the context of the estimation of the valuation of travel time savings (VTTS). The chapter pays particular attention to a number of key issues of interest to the DfT, notably the size of travel time savings as well as other interactions such as with income and distance.

The subsequent discussion takes the 1999 study by Accent and Hague Consulting Group (AHCG) as being the default toolbox, with small additions made by the ITS/Bates work in 2003. The discussion looks at whether the 1999 methods may have partly influenced the findings or limited the scope for analysis, and also looks at how design and estimation methodologies have evolved since 1999 and how these developments may come in useful in any new studies. This is especially important given the range of new policy questions that a new VTTS study is likely to focus on. Finally, a number of crucial issues that need to be addressed by any new study are also discussed. At the end of both the design and the estimation section, a table is included to compare the approaches used in the AHCG study to the current state of practice, international evidence and the state of the art. Reflecting the emphasis of the 1999 study, the focus of this chapter will be on analysis of the car mode, although many of issues covered are equally relevant to freight.

To some extent, especially in the estimation work, the AHCG study was ahead of its time, a point that we will return to below. Nevertheless, survey and estimation capabilities have evolved very significantly over the last fifteen years, and while the benefits of relying on tried and tested methods, partly with a view to facilitating comparisons, should not be overlooked, the potential advantages of new methods should at the very least be taken into consideration when designing an update study. Indeed, in our view, the fact that methods have progressed so much is in itself a reason for a VTTS update study.

Already at this early stage of the discussion, it is worth stressing that while advanced approaches may involve a level of detail and complexity that is not appropriate for all application contexts, results from such approaches can be simplified while at the same time still benefiting from any bias reductions achieved by these methods. As an illustration, while the end user of a VTTS study may be solely interested in the mean or median valuation, the use of advanced estimation approaches that allow for complex heterogeneity patterns in these valuations across respondents may permit the derivation of a potentially more unbiased mean/median measure .

As mentioned above, both survey and estimation methodologies have evolved significantly in recent years. At the estimation stage, it is only natural to attempt to make use of the recent gains in modelling flexibility and to compare the results to those obtained with more basic approaches. This process clearly involves no risk, comes at little additional cost, but has potentially significant rewards. The question of which survey method should be employed is perhaps more contentious, however. The actual survey methods used in the 1999 AHCG work have since been extensively reused in other national studies, including in the Netherlands and Denmark, and there will undoubtedly be a feeling in some quarters that a new study should rely once again on these tried and tested methods. This is a crucial decision and should not be taken without careful consideration. Without any preconceptions, this chapter will discuss the advantages and disadvantages of continued reliance on this design approach, and its potential influence on results. This is especially important given the limited success of the 1999 study in dealing with gains and losses and small time savings (STS).

3.2 Survey issues

This section discusses the issue of how data is collected for use in studies looking at the estimation of VTTS measures. We will begin by reminding ourselves of the rationale for SP surveys for VTTS, before considering the basic format of such experiments - especially in terms of the choices/alternatives that are being offered to survey respondents.

The discussion will then proceed to more detailed matters. First, we will consider the important question of how best to express the cost attribute, central of course to the formula for VTTS. Second, we will address a series of design features, namely the numbers of alternatives and attributes, the number of attribute levels, the values for these levels, and finally the underlying statistical design; we also introduce the issue of „referencing' (section 3.2.4), which is relevant to STS. Finally we address some practical issues of implementation, including sampling strategies and survey administration.

3.2.1 Survey technique

While in 1999, the case for using SP still had to be defended (see e.g. second paragraph on page 139 of the AHCG report), this method is now firmly established as the main approach for VTTS studies in the United Kingdom as well as many other countries, including The Netherlands, Switzerland, Denmark, Australia, the United States and Chile, to name but a few.

Although SP surveys are in general well suited for use in VTTS studies, there is little or no guidance on the appropriateness of using them when looking at valuations of STS, a topic that was of interest in the 1999 study and which has been specifically identified as one of the DfT's areas of concern for the current study. Here, the use of SP potentially needs to be reviewed, as discussed later in this chapter.

3.2.2 Context

A crucial first decision that needs to be taken in the development of a SP survey is the context of the choice situations. The three most common examples are mode choice, route choice, and the choice of departure time. In some cases, a combination of more than one of these contexts is used within a given choice set (e.g. combined choice of mode and departure time), while many studies also present respondents with a range of choices covering the different contexts. With reference to our earlier comment under Option A1 in Chapter 1, we note that the 1987 and 1999 studies involved a number of sub-tasks/SPs, encompassing various such contexts. The scope of the 1987 study was broader than the 1999 study, and encompassed both within-mode (covering car and public transport) and mode choice surveys. The 1999 study was solely within-mode, and passenger surveys were restricted to the car mode; the subsequent discussion will describe these surveys in more detail.

The 1999 study involved several sub-tasks, specifically:

- An abstract value of time exercise, giving respondents a choice between two different untolled routes, with one route being faster while the other is cheaper
- A set of three separate SP exercises not involving travel cost, with each respondent receiving only one of the three, where the three SPs look at binary choices encouraging respondents to trade between travel time and:
    - road characteristics
    - departure time
    - chance of delay
- An exercise involving the choice between a tolled current route and an untolled alternative route
- A rerun of the 1980 Tyne crossing SP, giving a choice between the Tyne Bridge and the Tyne Tunnel, with alternatives described by petrol cost, toll cost (for tunnel only), total travel time, and the two subcomponents of free flow and slowed down travel time
- A new Tyne crossing SP, again involving a choice between the Tyne Bridge and the Tyne Tunnel, but with alternatives described solely by travel time and travel cost, which includes a toll for the Tyne Tunnel

The recommendation from the AHCG study was to use the abstract experiment for appraisal reasons, with the toll road experiments only being used in toll road contexts. The rationale behind this recommendation is that the abstract scenario is a more controlled setting while also providing the most general context. Additionally, this experiment is clearly not affected by the political voting observed in the toll road studies (see later discussions). Nevertheless, it should be noted that a number of issues possibly arise with the abstract approach. The main one of these is the cost attribute, which we will return to in the next subsection. An additional criticism could be levelled at the design in terms of being overly simplistic, an issue which we will similarly return to later in this chapter.

3.2.3 Cost attributes

Independently of the context of a survey, the main emphasis is on using the hypothetical choice scenarios to study the relative sensitivity of respondents to time and money components. The latter especially deserves some special attention. While in a public transport context, the cost component of the journey is relatively easily understood (i.e. journey fare), complications arise for car journeys. Indeed, the main cost component of a car journey is running cost which, although some weight should also be given to maintenance costs and depreciation, is essentially the fuel cost for the journey. Many studies rely extensively on running costs, but it is important to recognise that this is a difficult concept for respondents to comprehend, not least because fuel bills are not generally paid on a journey by journey basis.

Since analyses of VTTS focus upon the trade-off between time and money, there is a natural temptation to ’design in’ negative correlation between travel times and costs in the SP experiment (i.e. implying that a faster journey is more expensive than a slower journey). This generally makes sense in a public transport context, but can create problems in the car context, since in real life scenarios, travel time and running costs are strongly positively correlated.

David Hensher, one of the leading advocates for realism in SP design, has put forward the notion of "*experientially meaningful configurations*", i.e. ensuring that respondents are presented with choices that would be representative, at least up to a degree, of real life scenarios so as to ensure an acceptable degree of realism and response quality. This is arguably not the case in experiments based on negative correlation between travel time and running costs, and exposes analysts to the accusation that such experiments may induce undesirable influences on the results.

Given the difficulties of relying solely on running costs, there is considerable interest in exploring the use of other cost components, especially road tolls. In many ways, toll road studies offer one of the most realistic settings for studying valuations of travel time reductions, with a higher toll (or indeed a non-zero toll) applying to more rapid routes. This could be in the form of comparing several tolled routes, with negative correlation between the level of toll and the travel time (i.e. a more highly tolled route is faster), or a comparison between slower untolled routes and faster tolled routes.

Two main complications also arise in toll road studies, however. The first of these is lack of experience, which is especially relevant in a UK context. Indeed, with the exception of a few tolled river crossings and the M6 toll road, the UK is relatively toll free (congestion charging is a separate topic) and a large share of the driving population will have little or no experience of toll roads. This not only causes problems with sampling strategies, which we will return to below, but also potentially means that some respondents of the survey may have difficulties relating to the *toll* attribute.

The lack of experience with toll roads arguably also accentuates the second problem with such studies, namely that of a high aversion to tolled options. Road tolls are a contentious issue, and surveys including a toll attribute are often affected by a high level of political voting or lexicographic behaviour, with respondents refusing to choose a tolled option (if untolled options are available) or always choosing the option with the lowest toll. This is especially the case in surveys making use of respondents with limited or no exposure to the benefits of toll roads, and/or respondents currently using an untolled route, such as the majority of the UK population. Not surprisingly, results in such UK studies are often affected by misunderstanding as well as strategic bias, as recently observed by Chintakayala et al. (2009).

In other areas, such as notably in Australia and some areas of the US, where toll roads are quite common, VTTS studies regularly make use of toll road settings. In summary, while toll road choice scenarios offer more "experientially meaningful configurations", they also lead to a number of additional issues.

As the above discussion shows, both the use of running costs as well as the use of toll costs can cause complications, of a different nature, and for different reasons. An alternative comes in using surveys and modelling to obtain non-monetary valuations for attributes of interest and to monetise them on the basis of existing valuations. Here, the AHCG study for example included choices between alternatives with different road conditions, inviting respondents to trade between travel time and road conditions. Other options include trading between travel time and reliability, or different components of travel time (e.g. free flow vs slowed down). The problem with approaches of this type is that, in order to monetise the results, a valuation for one of the attributes needs to be obtained from elsewhere. As an example, we may use an existing monetary valuation of road conditions in order to infer the VTTS on the basis of the estimated trade-off between road conditions and travel time. Such an approach can be advantageous in cases where it is easier to obtain a monetary valuation for one of the trade-offs than is the case for the other one, for example due to political voting. In other words, a respondent may be more likely to reveal his/her true relative sensitivities to travel time and schedule delay in an experiment not involving cost.

As already mentioned above, the main emphasis in the AHCG work was on the abstract choice scenarios. In the absence of a road toll, it has to be assumed that the cost component is primarily the running cost, although this is not explicitly stated in the surveys. In the toll road studies used in the AHCG work, running costs were included on top of toll costs, though without variation in the abstract toll road experiments. In both the abstract toll road experiment and the Tyne Crossing study, choice situations are included in which a faster alternative has a higher (or equal) running cost than the slower competing alternative, i.e. using the same rationale as in the abstract scenarios.

3.2.4 Design

The design of a SP survey has a number of separate components. The analyst needs to first make decisions on the number of alternatives, the number of attributes, the number (and rate of use) of levels, the values for these levels, and finally the underlying statistical design. Special care is also needed if issues such as referencing and STS are of interest. These components will be the subject of the following discussion.

A choice situation in a SP survey presents a respondent with a fixed number of mutually exclusive alternatives, each described by a number of attributes. In generating a design for a survey, the analyst first needs to decide on the number of alternatives and attributes.

Alternatives and attributes

The majority of the weight in the AHCG analysis and in the generation of VTTS measures is given to the abstract choice between two untolled alternatives, described by travel time and travel cost. Even the remaining choice scenarios are based on a very limited number of attributes. Such simplistic choice scenarios clearly avoid any risk of overburdening respondents, and this, in conjunction with the use of paper based surveys, was the main motivation for such an approach. However, work has not only shown that respondents can adequately deal with a larger number of alternatives, but that unnecessarily restricting the number of alternatives may in fact make the surveys too simplistic and transparent, while also bearing little resemblance to real life scenarios. For example, on the basis of a proposition that the scale of the random error term proxies for the reliability of the data, Caussade et al. (2005) recommend four as the optimal number of alternatives. Additionally, a case can be made for increasing the number of alternatives on the grounds that this allows for greater variability in each choice set, increasing data richness while also reducing the overall sample size requirements.

The incorporation of other attributes into the choice situations, such as departure time, reliability, or different travel time components, may also be advantageous for three reasons. Firstly, they lead to a higher degree of realism, potentially improving response quality. Secondly, they mask the aim of the study, possibly reducing the risk of political voting or self selection. Finally, they obviously allow for the study of valuations in a broader context. As mentioned above, the main argument against increasing the complexity of stated choice scenarios is that of respondent burden. However, it has now been suggested that not only are respondents able to cope with relatively complex scenarios (see e.g. Caussade et al., 2005; Chintakayala et al., 2009), but that making choice sets relevant by including all important information may in fact improve response quality (see e.g. Hensher, 2006). Furthermore, from policy and analytical perspectives, there is an interest, particularly in public transport, to value as many explicit attributes as possible, thereby providing insights on the factors influencing travel behaviour and „unpacking' the alternative-specific constants typically specified in models.

Incorporating reliability

An additional attribute that is likely to be of interest in any new study, and which was included in one of the AHCG surveys, is travel time reliability. This topic is now of such importance that it is regularly looked at in VTTS studies, such as for example in the recent Dutch scoping study. A major issue in this context is the presentation of travel time reliability, where different methods exist, ranging from simplistic travel time variability approaches (e.g. +/- 5 minutes), to descriptive or graphical approaches showing the likelihood and extent of a delay. The jury is still out on which of these approaches is most adequate. Additionally, the interactions between travel time and travel time reliability need careful consideration, partly with a view to avoiding a situation where one attribute dwarfs the other in importance, leading to lexicographic behaviour. In this context, we point the reader towards the recent DfT Rail study (Batley et al., 2008).

Levels

Another important decision relates to the number of levels used for each attribute in the design, and the actual values for these levels. Here, the main emphasis is generally on using a set of levels that is broad enough to allow for a diverse set of possible VTTS measures while also not being so wide as to lead to unrealistic combinations.

In some practical studies, relatively little attention is devoted to the number of levels and the number of times each level is used in the survey. Raising the number of levels increases the number of possible combinations, improving the richness of the data, up to a point where the effects become detrimental (see e.g. Chintakayala et al., 2009). It is however similarly important to ensure an even representation of the levels. In fact, work at the Sawtooth software company in the 1980s showed that an uneven representation of levels may lead to respondents focussing more on one attribute than on other attributes. Here, it should be noted that in the design used in the AHCG study, there were five levels for each attributes, including the current level, where each attribute was used once per respondent except for the current level which was used four times. The fact that the same approach was used for both time and cost should at least avoid the above issue of uneven representation across attributes, but the uneven representation within attributes arguably reduces the richness of the data while also potentially drawing more attention towards the current level of an attribute.

Statistical design

Notwithstanding recent developments, which we return to below, the majority of SP questionnaires are still based on orthogonal designs. In an orthogonal design, the different columns in the design are uncorrelated, allowing all main effects to be indentified from the data. However, the use of orthogonal designs also poses a number of complications, primarily to do with dominance. Looking at a simple binary choice scenario with two attributes (travel time and travel cost) and two levels each (15 minutes and 30 minutes; £1 and £2), a full factorial orthogonal design would be given by:

Table 3.1: Binary choice design characterised by dominance

Time A	***Cost***	***Time B***	***Cost***
15 minutes	£1	15 minutes	£1
15 minutes	£1	15 minutes	£2
15 minutes	£1	30 minutes	£1
15 minutes	£1	30 minutes	£2
15 minutes	£2	15 minutes	£1
15 minutes	£2	15 minutes	£2
15 minutes	£2	30 minutes	£1
15 minutes	£2	30 minutes	£2
30 minutes	£1	15 minutes	£1
30 minutes	£1	15 minutes	£2
30 minutes	£1	30 minutes	£1
30 minutes	£1	30 minutes	£2
30 minutes	£2	15 minutes	£1
30 minutes	£2	15 minutes	£2
30 minutes	£2	30 minutes	£1
30 minutes	£2	30 minutes	£2

Here, it becomes immediately clear that a large number of choice situations in this design are affected by dominance, with one alternative being both cheaper and faster, while four choice situations also present a choice between two equal alternatives. Some studies overlook this issue, and retain such choice situations in the design, with the outcome that dominant choices add nothing to our understanding of choice processes and potentially have detrimental effects on response quality. Other studies take a more aggressive approach, simply removing these problematic choice situations. As an example, the recent Dutch scoping study states:

> *"...[given] the low number of dominant questions produced by the orthogonal design, we therefore used a standard orthogonal design (eliminating all dominant questions) for this experiment"*

A problem with this approach is that it often leads to a loss of orthogonality, and almost invariably also leads to a loss of „attribute level balance'.

In the AHCG work, this issue was avoided in the abstract choice scenario by making use of a „manual' design, whereby the analyst intervenes to eliminate any dominance that might be prevalent in an orthogonal design. Specifically, each respondent was faced with eight choice sets, relating to the respondent's current trip. In four of these eight choice sets, the current trip was included as one of the alternatives, with two of them involving a comparison with a faster but more expensive option, and two of them involving a comparison with a slower but cheaper option. In the remaining four choice sets, the current time was used for one of the alternatives, with the current cost being used for the other alternative. Each time, the levels for the remaining alternative were chosen so as to encourage trading. As an example, if option A uses the current time and option B uses the current cost, then if a cost reduction is used for option A, a time reduction would be used for option B. An example is given below, using the abstract design from the first questionnaire.

Table 3.2: Example ‚manual' design

time a	cost a	time b	cost b
Current	Current	- 3 mins	+ 15 p
Current	+ 125 p	+ 5 mins	current
Current	- 35 p	- 5 mins	current
+ 10 mins	Current	Current	+ 35 p
- 5 mins	+ 75 p	Current	current
- 3 mins	Current	Current	- 5 p
Current	Current	+ 10 mins	- 10 p
+ 5 mins	- 50 p	Current	current

It can be seen straightaway that this design avoids issues of dominance, which is clearly desirable. This is one of the reasons for the popularity of this design (named after its developer Mark Bradley), as acknowledged by the recent Dutch scoping study:

> *"For the first experiment, we used a so-called „Bradley design' to generate the choice sets for the underlying design. The merit of the Bradley design is that it does not produce any dominant choice set in the SP questions."*

While the properties in terms of dominance are very desirable, the Bradley design as used in the AHCG study also has a number of other properties, the effects of which need to be discussed. Looking first at the correlation structure in the design, we can see negative correlation between time and cost attributes for the same alternative. In other words, if the cost of an alternative increases, its time is reduced. This encourages trading between time and cost, but potentially causes issues with realism, as discussed previously. There is positive correlation between the time for alternative A and the cost for alternative B, with the same applying for the cost for alternative A and the time for alternative B (Table 3.3).

If one alternative is faster, the other is cheaper, and vice versa. However, with the specific attribute levels chosen by the AHCG team, the correlation levels are not uniform. Additionally, the costs are different for the two alternatives, and there is a lack of attribute level balance. All these factors potentially have an influence on behaviour that should ideally be tested before reusing a design of this type.

Table 3.3: Correlations between variables in Table 3.2

	time a	*cost a*	*time b*	*cost b*
time a	1			
cost a	-0.38429	1		
time b	-0.04007	0.370391	1	
cost b	0.717191	-0.08737	-0.37669	1

Whilst highlighting such features of the Bradley design, it is important not to overlook the fact that (in some shape or form) it has been applied not only in the previous UK national study, but also the national studies of the Netherlands, Denmark, Sweden and Norway. The continued use of the Bradley design reflects an opinion shared by many analysts that it has - in the main - delivered successful results.

In the course of these applications, the manner in which the Bradley design has been applied has evolved over time. In the recent Danish study, for example, the use of the Bradley design in combination with models in logWTP space brought some attractive functionality, not least the ability to directly plot the VTTS distribution.

Furthermore, the levels for the Bradley 2x2 design can be varied in a manner that is convenient for the estimation of VTTS (indeed this approach was followed in Denmark, Norway and Sweden), as follows:

- Design is referenced against a recent trip, the time and cost of which is agreed with the respondent.
- Two time differences are drawn using stratification from some predefined set; changes should be meaningful both up and down from the reference.
- These are then applied to all four (difference) quadrants, such that there are eight choices in total.
- Trade-off VTTS are then drawn with stratification from some distribution, which should (almost) cover the VTTS distribution.
- Time difference multiplied by trade-off VTTS gives cost difference.
- Signs of differences from the references are inferred from the relevant quadrant.
- Levels are obtained by adding the reference.
- Denmark added a dominated choice as a check. Norway and Sweden added an off-reference choice instead. Ex post, it was judged that Denmark used too low a value for the trade-off VTTS, but this was corrected in Norway and Sweden.

Finally, some concerns have been recently raised about respondents' ability to adequately distinguish between the alternatives in the survey (Hess et al., 2009), where this was based on the Danish data, but where similarly low probability of correct prediction was observed in the UK and Dutch data.

The field of experimental design has witnessed important changes since the time of the AHCG work. Aside from work dealing with respondent burden and survey realism, a majority of the effort has gone into improving the efficiency. Given the expensive nature of data collection, there is great interest in reducing the number of individual respondents used in a given study. One way of achieving this is to increase the number of observations per respondent, hence obtaining a larger sample with fewer respondents, but this comes at the cost of heightened respondent burden with likely detrimental effects on data quality. The aim is thus to reduce the number of respondents without unnecessarily increasing respondent burden.

In recent years, the statistical state of the art of designing SP experiments has moved away from orthogonal designs to efficient designs. Various methods exist, but the basic rationale is the same. A statistically efficient design is a design that minimizes the elements of the asymptotic (co)variance matrix with the aim of producing greater reliability in the parameter estimates given a fixed number of choice observations. That is, a statistically efficient design is constructed with the aim of maximizing the asymptotic t-ratios obtained from data collected using the design. The net outcome is that the same level of statistical robustness can be achieved with a smaller sample.

It is worth observing that whilst these more *advanced* design approaches have found only occasional use in the UK, they have been widely adopted elsewhere, notably in Australia. Specialist software now exists that allows the generation of efficient designs, and while the time investment at the design stage is slightly higher than for orthogonal designs, this can be easily offset by savings in data collection cost resulting from smaller sample sizes. Another point in favour of these advanced design techniques is that they can be easily adapted to incorporate constraints for avoiding dominance, while also being well suited for work looking at gains and losses, as discussed below.

Against the attractions of efficient designs, one must however acknowledge their weaknesses and complexities. A fundamental problem is that the construction of an efficient design requires the analyst to establish the „likely' asymptotic (co)variance matrix for the design prior to collecting any data, implying that the attribute levels as well as likely parameter estimates are known in advance. In practice, this can be addressed by using an orthogonal design as a starting point and producing prior values for the main survey wave through a pilot survey. That said, we have no insight as to the „true' model a priori. In particular, we have no insight as to the shape of the VTTS distribution. A design that is very good at recovering the parameters of, for example, a lognormal distribution may be poor at distinguishing lognormal from other distributions. This is because the design relies on the assumption that the distribution is lognormal; this is the source of efficiency.

Indeed, a design that seeks to identify the shape of a distribution may be quite different from a design that seeks to identify the parameters of a certain model. Ideally, we would wish to estimate the VTTS distribution. A choice observation basically informs us (after noise) whether the VTTS of a respondent is smaller or larger than the trade-off implicit in the choice. Given many observations of choice under a certain trade-off, we can estimate the share of respondents who have VTTS smaller/larger than the implied trade-off. This implies a need for many different trade-offs, extending across the VTTS exhibited by different respondents. In this way, we outline a case for identifying the VTTS *distribution* from data, as opposed to focussing the design around specific „target' *values*.

“Referencing” and small changes

Given the interest of the present study in STS, there is a clear attraction in framing the survey around specific current levels. This implies a need for customisation, which was not a strong feature (aside from customisation to broad segmentations) of the 1999 study. One approach in this context is to include a reference trip as one of the alternatives in the survey, typically alongside two further hypothetical alternatives. This is for example the standard approach in many Australian studies. While this has the advantage of putting a reference trip *in front* of the respondent, it also potentially leads to large levels of inertia (in the form of non-trading) and special care is required during the design and the analysis.

While this approach has recently been shown to be of great use in analysing the differences between gains and losses (see e.g. Hess et al., 2008), its appropriateness was questioned by ITS/Bates (2003), who suggest that it may be preferable not to include such comparisons to the current journey in the design. The discussion by ITS/Bates relates not to pure pivot designs (in which the reference alternative is always included as one option) but to the AHCG work, in which a slightly different approach was used, with the current time and cost levels being used in each choice situation, but not necessarily for the same alternative. Indeed, two out of the eight choice sets for each respondent involved a choice between the current trip and a faster but more expensive trip, two situations involved a choice between the current trip and a slower but cheaper trip, and the remaining four scenarios involved departures from the current alternative along one dimension for each of the two alternatives.

While this specific combination of choice sets may be desirable from an estimation perspective, allowing for a comparison of different departures from the reference alternative, the main issue is whether the design is well suited to study the differential response to gains and losses. As highlighted above, the current alternative is included in half of all choice situations. This is midway between including the reference alternative in all choice sets (i.e. as in the typical Australian approach) and basing the design on a reference trip but not including the actual reference trip in the survey. No discussion is given as to which approach is better, and this is an important area for further work. Using the reference alternative may facilitate reference point formation, but may also lead to higher levels of inertia.

ITS/Bates (2003) show that when working with the AHCG data based on the Bradley design, the retrieved difference between gains and losses is in fact an artefact of the model specification and that a statistically superior model actually shows no differences. On the other hand, work by Hess et al. (2008) on a more traditional pivot style dataset shows clear differences between gains and losses even after accounting for inertia. This poses the question of whether a more traditional pivot design may be better suited for analysing the difference between gains and losses. Overall however, it is unclear as to how the SP presentation may in fact influence the results in terms of gains and losses.

Independently of the nature of the design, another issue in this context is the actual definition of the reference point. Here, the question needs to be asked whether the *current trip* is actually the most natural reference point for an individual when it could equally well be the *ideal trip*. This issue needs to be kept in mind at the design stage (to enable adequate pivoting) but is also of crucial importance at the modelling end, as discussed later on in this chapter.

An added complication is that the reference point may be subject to endogeneity. Individuals choose the trips that they undertake (and they themselves report the characteristics of their trips), such that the characteristics of their chosen trip depend on their preferences. This means that „independent' variables (e.g. time and cost) of models are not in fact independent. This is a serious problem, which arguably deserves much greater attention in practical work. At present, there are perhaps three camps on this issue. One camp recognises the problem and chooses to design choice situations in an exogenous way. That is to say, customisation is sacrificed in order to avoid bias. The second camp recognises the problem, maintains an endogenous design, but tries to account for endogeneity through the econometric modelling (e.g. Train & Wilson, 2008). The third camp, which seemingly accounts for most transport researchers, simply does not recognise the endogeneity problem.

Another issue of great interest is the valuation of STS or in more general terms the difference in valuation between small and large savings. While the consistency of such differences with economic theory has been questioned, it is nevertheless important to ensure that a design has the potential to allow for such differences to manifest themselves in the data. The AHCG work reports evidence of size effects, with increasing valuations as the size of savings goes up; these results were however called into question by ITS/Bates (2003). Here it would be important to determine whether this is an effect of the design. In more general terms, it remains to be discussed whether SP is an appropriate tool for dealing with small time changes. Indeed, the three reviews for the original AHCG work all questioned the adequateness of SP for this purpose. In this context, the significant evidence of threshold formation by respondents in SP surveys (see e.g. Cantillo et al., 2006) is also of interest. In this context, ITS/Bates make the case for using RP for small changes, but it may be difficult in practice to find appropriate case studies. Another question that needs to be addressed in this context is whether it is the mix between small and large savings in SP studies that causes these effects, and whether focussing on different sizes of time savings in separate surveys may be preferable. Indeed, in a survey where a respondent is faced varyingly with large and small changes, the response to small changes may be understated. A related point is the recurrent practical question of whether we should concern ourselves with „one off' time savings that are in themselves small in magnitude, or the accumulation of such time savings over days, weeks or months. Whilst the latter would perhaps apply to regular commuters, one should remain mindful of the concomitant point that irregular travellers may have insufficient experience of the journey in order to distinguish a small time saving from background variability.

3.2.5 Sampling strategy

Two main points need to be addressed at the sampling stage, namely sample size and sample selection.

Decisions need to be taken in relation to the overall sample size as well as the number of respondents falling into various socio-demographic groups, if such quota are to be used. Many surveys use rather arbitrary approaches for sample size decisions, e.g. a round number of respondents (such as 500) in each quota group. This seems to be the approach taken in the AHCG study, although the sample size was admittedly very large.

Ad hoc formulae also exist to determine the minimum number of respondents required for identifiability of model parameters (see e.g. Orme, 1988), but over the last few years, significant headway has been made in developing formulae that take additional prior information into account to provide more accurate guidance on necessary sample sizes (see e.g. Rose & Bliemer, 2007).

In a study aimed at drawing conclusions on the VTTS for the overall population, a certain degree of representativeness needs to be achieved. This can however be very problematic, as also highlighted in the AHCG study. With the issue of experience being especially pertinent to VTTS for toll road studies, AHCG took special care to identify only those respondents potentially affected by road tolls. This potentially reduces issues of strategic bias, but the results are inevitably influenced by the past experiences of these specific respondents.

Respondents for the abstract choice experiments were sampled at petrol stations, where motorway service stations were avoided given the argument that this would under-sample short distance travellers (relative to NTS). Nevertheless, ITS/Bates (2003) demonstrated that long-distance travellers are oversampled in the AHCG data. Not only are there possible correlations with socio-demographic indicators in this regard, but there are grounds for concern that this could have affected the findings in terms of STS.

3.2.6 Survey administration (data collection)

For data collection, a paper based survey was used in the AHCG study, an approach which still remains in widespread use in applied UK work. Paper based surveys are significantly cheaper than interviewer assisted surveys, and also have some advantages at the sampling end. However, a number of disadvantages arise, on top of the obvious issue of coding mistakes (during the translation from paper to data files).

Paper based surveys face severe limitations in terms of the degree of customisation, where typically, separate surveys are only prepared for a small number of segments, e.g. long distance vs. short distance (four bands were used in the AHCG work) and road type (motorway, urban and trunk were used by AHCG)

. Another issue is that in a paper based survey relating options to current values, the respondents are required to carry out calculations themselves. As an example, an alternative may be described as:

Current travel time + 10 minutes

Current travel cost – 50 pence

In this context, two main issues arise. Firstly, there is obvious scope for numerical mistakes. This would mean that a respondent who makes a mistake in the calculations will base his or her choices on values that are different from those later on used in estimation. This equates to an error in the explanatory variables which can lead to reduced explanatory power or even biased coefficients (in the case of systematic mistakes). Secondly, although surveys collect information on current travel times and costs, there is a small risk that a respondent may end up using different *current travel time* and *current travel cost* values than those given in the initial questions, and which are used as the basis for model estimation. This once again leads to errors in the explanatory variables.

Given the generally accepted knowledge that respondents struggle with the notion of percentage changes (again highlighted in the AHCG work), absolute time changes have to be used, which may create problems if the number of separate surveys is low, i.e. the same savings are used for potentially quite different journey times.

All the above problems can to a large degree be avoided by making use of a computer based survey, either in the form of an interviewer assisted survey or an internet based survey. Such surveys allow for a high degree of customisation and also carry out calculations automatically, avoiding numerical issues while also guaranteeing a correspondence between those values used in the survey and those values used in the models. Furthermore, such surveys can rely on percentage variations, which may produce more realistic attribute levels, while also increasing data richness.

The cost of interviewer assisted surveys may sometimes be prohibitively high, pushing researchers towards internet based surveys. Here, issues of sample representativeness may be avoided by sampling respondents in tried and tested ways (e.g. roadside) and by handing out login details for the internet based survey, an approach that is again becoming more common place, for example in many river crossing surveys in the United States.

3.2.7 Survey testing

Independently of the decisions made in relation to survey design, it would be good practice to undertake careful testing, comparing various possible approaches, prior to the main survey. This would have to involve the collection of pilot data. Indeed, a point that is often not recognised is that while simulation can show whether a design is able to capture specific effects, the impact of the design on actual behaviour cannot be tested in this manner, an issue that is especially pertinent in the case of issues of realism. From this perspective, the quote from Fowkes and Wardman's review of AHCG that:

> *"... we subjected the [AHCG] designs to testing using synthetic data and found it to be able to recover a large range of values of time remarkably well. We are confident that the results being obtained are not artefacts of the design."*

should be taken to say that the results obtained by the models reflect only the behaviour observed in the sample, and it should not be presumed that this behaviour is not itself influenced by the design. On such matters, pilot surveys involving a comparison of different designs can help to provide insights.

A final point that needs addressing is the issue of testing for respondent rationality. It is becoming popular to include so called *„no brainer'* choices in surveys, generally in the form of dominated choices, and to eliminate any respondents failing these tests. While we recognise the importance of such tests, especially in the case of a departure from current methods, we feel it is important to note that such tests have to date often been performed in a potentially inappropriate manner. As an example, the Danish study included a dominated choice as the sixth (out of nine) choice sets. The problem with this approach is that, very much in the same way as retaining dominated choices in standard orthogonal designs, the presence of this choice scenario may lead to respondents not taking the remainder of the survey seriously. For this reason, it is our recommendation that if such tests are to be included, these should be done at the end of the survey, thereby avoiding any biasing influence on the remainder of the data.

3.2.8 Summary: survey work

The following table presents and compares the AHCG methods with current practice in the UK and abroad, as well as the state of the art.

Table 3.4: Summary of survey methods

Topic	**AHCG work**	**UK state of practice**	**International practice**	**State of the art**
Survey context	Main weight given to abstract route choice experiment Additional surveys looking at road characteristics, departure time, and chance of delay Three experiments involving road tolls		Similar reliance on abstract route choice in Netherlands and Denmark Toll road studies in countries familiar with tolls Mode choice experiments in some countries e.g. in Switzerland	Not altogether agreed; some experts advocate surveys closely related to actual circumstances, other experts advocate abstract designs
Car costs	Major reliance on running cost, with tolls only used in areas where currently in operation		Tolls used in many international studies	N/A
Alternatives and attributes	Binary experiments, mainly with two attributes only	Mainly binary experiments, but occasionally with more attributes	Often binary, with main exception being Australia and South America Generally more than two attributes	No evidence that respondents are overburdened by 3-4 alternatives; some evidence to suggest that more is better Strong emphasis on realism and relevance, with some surveys dealing successfully with large number of attributes
Attribute level balance	Design unbalanced, with current level used in 50% of choice sets, with four remaining levels used once each	Designs often unbalanced as a result of using manual designs or inappropriate software, with a few exceptions, mostly outside of Europe		Balanced designs, except in scenarios where choice sets are problematic
Statistical design	Manual Bradley design	Manual designs or orthogonal designs	Manual/orthogonal designs, with some reliance on efficient designs outside of Europe	Statistically efficient designs
Referencing	Attribute levels based on current experience, but not necessarily including current trip as one alternative		Similar to UK practice, though with reliance on pivot designs in some countries	Pivot designs, often including reference trip directly as one alternative
Small time savings	No special consideration is typically given, so that small and large savings are presented in the same survey			
Data collection	Paper based	Paper based, but increasingly also telephone and computer based	Still some paper based, but mainly computer based, and also internet based	Computer based, with high degree of customisation, either face to face, or internet based

3.3 Estimation work

This part of the chapter looks specifically at estimation methodology as opposed to design approaches. As already mentioned previously, the AHCG work was in many ways ahead of its time, containing innovations along multiple modelling dimensions. Nevertheless, the state-of-the-art and the state-of-practice have evolved significantly since 1999, and various new approaches are now at the analyst's disposal. As with the survey methodology discussion, the following is divided into a number of subsections looking at different estimation topics. This discussion is intentionally kept shorter than the design part. Indeed, it is the authors' opinion that any new study should consider the costs and benefits of applying the full set of available tools at the estimation stage; this involves essentially no additional risks, since basic methods are always available as a fallback option.

3.3.1 Pre-analysis processing

Non-traders

Many SP datasets are characterised by a large number of respondents who consistently choose the same alternative. This is especially true in the presence of reference alternatives, in mode choice scenarios, and in toll road experiments. The AHCG work experienced problems with non-trading in the toll road example, with up to 30% of respondents consistently choosing the untolled option. The presence of a large number of non-traders in such a context will often lead to downward bias in the VTTS, an issue that the AHCG work tries to address by removing these respondents (although this only leads to small changes in results). A possible reason for this is the presence of a similarly large number of quasi-non traders, e.g. respondents who choose the untolled option in all cases except one.

One school of thought is that the analyst should endeavour to prevent such behaviour manifesting in the first place, by taking appropriate steps in the survey design stage, e.g. avoiding wherever possible situations that might lead to strategic voting, and offering large enough incentives to encourage respondents to trade (e.g. Hess et al., 2009). Nevertheless, the question remains as to what to do in the presence of datasets affected by a large rate of non-trading.

It is important to note that simply removing non-traders is a very arbitrary step, affecting the representativeness of the data. Additionally, the question needs to be asked whether non-trading is in fact a reflection of real preferences or an effect of the study design. If the latter, the relevant respondents should most probably be removed as their non-trading behaviour should not be seen as being representative of their real world behaviour. However, if the behaviour is not a result of the survey design, their behaviour needs to be somehow accommodated in the models. To some extent, the use of constants can help mitigate the effect of non-traders and quasi-non traders, as seen in the success of using toll road constants in many studies (e.g. Hess et al., 2008). Here, a constant associated with toll roads would capture some of the political voting against toll roads that would otherwise have a potentially biasing effect on model estimates. However, even with this method, there is no guarantee of success, as was the case in the AHCG study.

Moreover, reference-dependence is a well established phenomenon in the literature (e.g. Tversky & Kahneman, 1991), which perhaps provokes a challenge to analysts to recognise that people exhibit this behaviour (at least in experimental and survey situations) and to develop methods to deal with it. If „consistency' is defined with respect to reference-free preferences, then it is quite clear that „inconsistencies' will occur under reference-dependent preferences. That said, the Bradley design is well equipped to detect such inconsistencies, as demonstrated in the De Borger & Fosgerau (2008) work.

To date, there is little consensus on how to distinguish between different types of non-traders, what their impact on model results may be, and how they should be dealt with, either in terms of removing them or accommodating them in the model. From this perspective, it is even more important to take special care at the survey design stage to reduce the risk of non-trading.

Consistency checks

As discussed in the design section, many studies now incorporate consistency checks in the surveys, and this provides analysts with some further scope for data cleaning. However, it is in this case crucial for the analyst to attempt to establish whether the inconsistent responses are due to mistakes by the respondents or are again an effect of the design. For this purpose, pilot studies are of great importance.

Outliers

Although good practice, it is very rare for choice modelling studies to engage in detailed outlier analyses. This is despite the widespread knowledge that even a very small number of outliers, or respondents whose behaviour does not conform with the model, may have an undue influence on model results. The presence of outliers may not only affect findings in terms of mean valuations but can play a very large role in the retrieved patterns of heterogeneity, for example artificially overstating the degree of heterogeneity. Such outlier analyses are in fact extremely easy to carry out and should be an essential part of any study. Once outliers have been identified in the data, they should not simply be removed from the sample; rather it is up to the analyst to determine the reason for the outliers and to decide whether attempts should be made to better accommodate them in the model.

3.3.2 Taste heterogeneity

Perhaps no single area in the field of discrete choice studies has evolved as much as the treatment of heterogeneity across respondents. A growing number of studies are based on random coefficients models, allowing the analyst to capture increasingly complicated patterns of heterogeneity that cannot be explained in a purely deterministic manner. These methods are also no longer the preserve of academic studies, as recently seen in the Danish VTTS study, which made use of some of the most advanced formulations in this context (similar methods were used in Norway and Sweden).

It is worth mentioning that the AHCG work made some initial first steps towards the treatment of random taste heterogeneity, with the use of mixed logit as well as latent class models. However, the state-of-practice and the state-of-the-art have evolved significantly, with a growing realisation that flexible specifications are required and that as few assumptions as possible should be made by the analyst in relation to the shape of the distribution of sensitivities in the sample population.

It should also be said that the growing interest in random coefficients models has unfortunately led to reduced efforts in linking taste heterogeneity to socio-demographics, when an understanding of the latter continues to be the priority for policy makers. We thus reiterate the point that analysts should always attempt to explain as much of the heterogeneity as possible in a deterministic manner prior to incorporating additional random variation. In this context, the latent class approach arguably has advantages over a simple mixed logit structure, by dividing the sample population into classes with distinct tastes where the probability of a respondent belonging to a certain class is a function of that respondent's socio-demographics. While the AHCG work did include such latent class models, the heterogeneity was in fact of a purely random nature and was not linked to socio-demographics.

Independently of the specific treatment of heterogeneity, an important issue needs to be raised. The main interest in many studies will simply be a mean measure of the VTTS, and this may lead to an opinion that the increased modelling effort required to accommodate complex patterns of heterogeneity is not justified. Here, it is important to note that not incorporating such heterogeneity in the case of often highly asymmetric distributions may in fact lead to biased mean valuations. From this perspective, even if the interest is solely in mean valuations, such valuations are significantly more reliable when obtained from models that recognise the presence of heterogeneity in the sample population. Given that it is generally impossible to explain a satisfactorily large share of heterogeneity solely through deterministic means, there is thus significant scope for random coefficients approaches even in very applied and policy-driven work.

3.3.3 Interactions and non-linearities

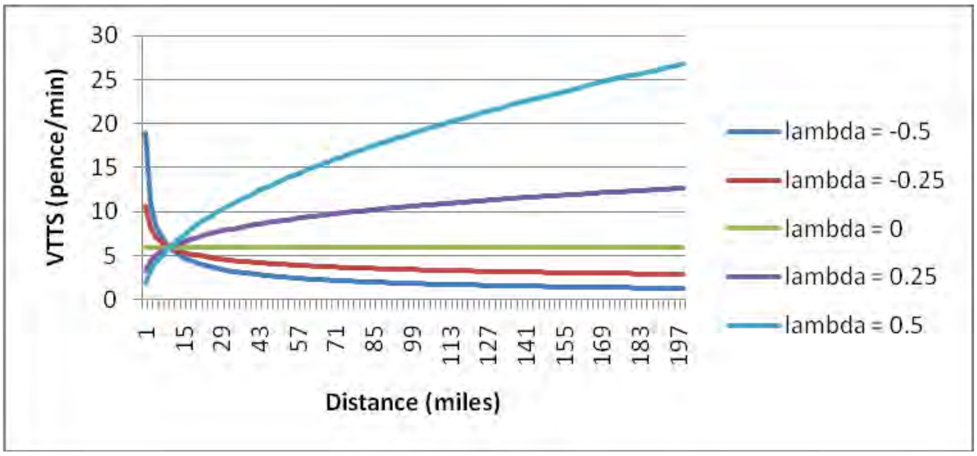
There is clear evidence, notably in the AHCG work, of non-linearities in response to changes in time and cost. At the same time however, there has always been the view that such non-linearities are not consistent with economic theory and that especially the sensitivity to cost should stay constant, independently of the attribute base level and the size of the change. This view was expressed for example in detail in the Glaister review of the AHCG work. While the arguments are entirely reasonable from an economics perspective, there is a large body of evidence to the contrary, sometimes leading to an apparent clash between economic and „behavioural plausibility'. The Bates review of AHCG states that while the sensitivities should be linear for CBA, the evidence from SP is very different, with Bates remarking "... it is quite clear that in practice individuals apply some concept of relativity". The issue is not just limited to the notion of decreasing or increasing marginal returns, but also extends to interactions between sensitivities and socio-demographic indicators. While an interaction between income and cost sensitivity is easily justified, many studies have similarly revealed an interaction between journey distance and cost and time sensitivities (see e.g. Axhausen et al., 2008), meaning that once again, a respondent's sensitivities, especially to cost, may not be stable across circumstances.

It is outside the scope of this chapter to address this issue in detail. However, with the clear evidence of significant patterns of non-linearity and interactions with socio-demographics in many datasets, some guidance is required on how these phenomena should be dealt with at the estimation stage.

The AHCG work allowed for interactions on time and costs, but to a large extent relied on linear specifications. Differential responses by size were accommodated through the estimation of separate coefficients in different bands. It should be clear that linear specifications and simple segmentation approaches reflect a much simplified representation of reality, and that sensitivities are not likely to either change linearly or in steps. For this purpose, the use of continuous specifications may be preferable. For non-linearities, the standard approach is to use a log transform to allow for decreasing marginal returns. This however again amounts to a strong assumption, and it may be preferable to make use of a more flexible specification, such as Box-Cox, that allows the degree of non-linearity to manifest itself from the data rather than being postulated by the analyst. Importantly, this also still allows for increasing marginal returns.

In terms of interactions with socio-demographics such as income and trip specific factors such as distance, the ITS/Bates work put forward a flexible continuous elasticity specification, allowing the analyst to directly estimate for example the cost elasticity in relation to income or distance. Using the example of distance, this specification is based on replacing $\beta \cdot x$ (where x is an attribute such as cost and β is the estimated marginal utility) by $\beta \cdot (d/D)^{\lambda} \cdot x$, where d is the distance for the current observation, and D is the average distance. The estimate of λ gives the cost elasticity in relation to distance. While this specification is highly flexible and has now been reused for example in the Swiss VTTS study, it can also lead to complications when looking at very small changes, where the sensitivities may be overstated (in the case of decreasing marginal returns, i.e. negative λ) or understated (in the case of increasing marginal returns, i.e. positive λ. This is illustrated in the following example plot, where we use a base VTTS of 6 pence per minute, at a base distance of 10 miles.

Figure 3.1: Example plot of VTTS against distance



ITS/Bates acknowledge the general feeling that distance cannot in itself explain the variation in VTTS. However, the interaction with distance has been used extensively, for example also in the Swiss VTTS study, and ITS/Bates highlight this as an important avenue for future work. A related question is whether the interaction should in fact be with travel time instead of distance, given that the two are often not perfectly correlated. Finally, while ITS/Bates work solely with elasticities for the cost sensitivity, the Swiss work actually found that both time and cost sensitivities change with distance, and also with income. It should also be noted that the actual sign of the elasticities in relation to distance is surrounded by some uncertainty. ITS/Bates, working with the AHCG data, find that cost sensitivity decreases with distance but time sensitivity stays constant. By contrast, the Swiss work finds both sensitivities decreasing with distance, albeit at a faster rate for cost. Moreover, the net outcome of increasing VTTS with distance remains the same in both studies, and this may provide a justification for developing models in WTP space.

Finally, independently of the exact specification used, the importance of allowing for interactions and non-linearities in estimation should not be underestimated. Indeed, there is evidence (see e.g. Pinjari & Bhat, 2006) that a large degree of the unexplained heterogeneity in sensitivities in models is in fact the direct result of unmodelled non-linearities.

3.3.4 Computation of WTP measures

In VTTS stated choice experiments, respondents will be presented with alternatives described by a number of attributes that will include travel time and travel cost. In a typical modelling approach, respondents' sensitivities to these attributes will be estimated separately. On the other hand, the main interest from a policy perspective will be in the relative sensitivities to these two components, reflected in the valuation of travel time savings. In the most simplistic scenarios, where a single coefficient is estimated for each of the two components, this valuation can be calculated straightforwardly as the ratio between these two coefficients. However, major complications arise in the presence of random taste heterogeneity, where this valuation would now be obtained as the ratio of two random coefficients, where the properties of this ratio may be ill defined.

From this perspective, it can be preferable to make use of an estimation approach in which the VTTS is obtained directly as a model output. The AHCG study in fact made use of such an approach in what was arguably ground breaking work, although it did not explicitly refer to this with the now commonly used terminology of modelling in „WTP space'. With the likely reliance on advanced models allowing for random taste heterogeneity, any new study would benefit significantly from making use of such an approach, as reflected for example in the recent Danish VTTS study. In fact, it should be noted that the advantages of this approach are not limited to a random heterogeneity context, but that it may also be advantageous in the case of deterministic specifications, where the analyst is then able to study the link between socio-demographic indicators (such as income) and the valuation of travel time savings directly, rather than by doing so separately for travel time and travel cost sensitivities.

3.3.5 Gains and losses

The treatment of gains and losses depends on the specification of a reference point. This in turn depends in a large part, though not exclusively, on the survey used to collect the data. If a reference alternative is included explicitly in the choice set, then this will typically serve as the reference point in the models. The situation is made more complicated in the case where the reference trip influences the design, but is not included in all (as in the AHCG work) or any choice sets. In both of these cases, there is a degree of analytical uncertainty, the effects of which are not generally known. Some of these concerns were highlighted by the Glaister review of AHCG, raising the issue of a reference point that is not stable over time. While this may be less relevant in a SP context (where we are using an instantaneous panel), it might well play a role in actual choices that have lasting impacts.

The findings by ITS/Bates should also be kept in mind. In their re-analysis of the AHCG data, the inclusion of dummy variables linked to the current trip removed the asymmetry from the model results. By contrast, Hess et al. (2008) found no such effect when modelling a dataset in which the reference alternative was included in all choice sets (a situation that perhaps encourages the respondent to frame the choice).

3.3.6 Small savings and thresholds

As discussed in the design section, the ability to test for the valuation of STS depends to a large extent on the way in which small savings were presented in the survey. Additionally, the question remains as to the adequateness of SP as a tool for obtaining valuations for small savings. Here, there is a large body of work on threshold formation (see e.g. Cantillo et al., 2006), where this may be seen as a SP artefact. Another relevant area is the literature on Prospect Theory and reference dependent preferences generally; see De Borger & Fosgerau (2008) for an application to discrete choice models.

How to actually allow for different valuations for small savings is a separate issue. The specification of time bands has a potential influence on results, so that different sizes of bands need to be tested. On the other hand, continuous approaches may also not be well suited, given the earlier observations in relation to the ITS/Bates specification. Another issue is that the size of savings should arguably be linked to the overall duration. As such, 5 minutes on a short trip may carry a higher valuation than on a long trip. To some extent, this can be tested by appropriate interaction specifications, and relates to the question of whether respondents process VTTS choices in terms of absolute or relative savings.

One argument supporting the finding of very low valuations for small savings is that respondents do not see time savings as „bankable', and that a single minute each day has little or no additional benefit. Additionally, respondents may question whether changes would apply to all trips, and whether trip frequency might be relevant.

3.3.7 Software

The AHCG study relied on the use of the in house ALogit software for estimation work, with the possible exception of the latent class models. ALogit remains a very powerful estimation tool, and has a number of facilities that are extremely useful in applied work, most notably in terms of forecasting capabilities. Nevertheless, it should also be noted that various other packages exist that avoid some of the limitations of ALogit in terms of model specification. This especially concerns Biogeme, a freeware package that is now the most widely used tool for discrete choice analysis, and which allows the analyst to make use of almost all available state-of-the-art models. Any new study should be required to justify the use of a specific software package, and it is quite conceivable that a mixture of packages would have to be used to address all requirements.

3.3.8 Summary: estimation work

This section has looked at a number of different estimation topics, but three of these, namely the treatment of taste heterogeneity, the approach to interactions and non-linearities, and the computation of WTP measures, highlight important differences between the state of practice and the state of the art, as illustrated in the following table.

Table 3.5: Summary of estimation methods

Topic	AHCG work	UK state of practice	International practice	State of the art
Taste heterogeneity	Some initial steps towards Mixed Logit, and Latent Class, but majority of results based on models with linear interactions	Only limited use of Mixed Logit	Only limited use of Mixed Logit, but with some high profile studies making use of very advanced specification	Random coefficients models such as Mixed Logit with highly flexible distributions Latent class and discrete mixture models Non-parametric models
Interactions & non-linearities	Linear interactions	Often use log transforms, along with linear interactions, though some increasing reliance on more flexible continuous interactions		Continuous interactions, and non-linear transforms not making specific shape assumptions
Computation of WTP measures	Work in WTP space	Primarily work in preference space, with WTP computed as ratio, using simulation in the case of random coefficients		Work in WTP space

3.4 Areas of concern

This part of the chapter looks at a number of „areas of concern' identified by the Department, and discusses how these might be addressed methodologically in any new study.

3.4.1 *Further income segmentation vs. equity values*

While this is largely an appraisal issue, any reliance on greater segmentation at the *model use* stage clearly needs to be taken into account during model specification and estimation, as well as at the survey design stage (e.g. in terms of income bands). Greater segmentation does of course come at a certain cost in terms of data requirement. However, it is conceivable that this can at least in part be offset by making use of more efficient data collection approaches. Finally, as discussed in the estimation section, any non-linearities and heterogeneity present in the data are likely to have a biasing influence on mean values, if not properly dealt with. From this perspective, even if in appraisal only an equity value is to be used, it is of crucial importance to accommodate such heterogeneity at the modelling end.

3.4.2 *Group size/composition*

Variations in group size and composition may have an influence on WTP measures especially in a mode choice context where a distinction is made between modes in which a per passenger cost applies (i.e. most PT options) and cases where a per vehicle (e.g.

car/taxi) cost applies. If this issue is to be investigated in detail, then mode choice experiments may be an attractive approach. No significant complications arise at the modelling end other than dealing with inertia. Special care is obviously also required at the sampling stage.

3.4.3 Trip duration/distance

The effects of trip duration and distance have been discussed at length in the context of estimation. A number of questions arise. Firstly, it is not clear which of the two components is likely to be a better explanator of non-linearities. Secondly, the issue of how to model these interactions most adequately needs to be addressed. Here, a comprehensive case exists for the use of continuous specification rather than segmentation, but a number of issues arise, notably in terms of functional form, and in the presence of data collected in the form of distance bands. Thirdly, specific survey questionnaires may be required for different distance groups, and special care is also required in sampling.

3.4.4 Trip purpose

At the estimation stage, trip purpose is one of the most obvious variables to use in segmentation, and the differences across purposes is often sufficiently large to warrant the use of separate models.

3.4.5 Mode

Mode choice experiments can be a useful tool in VTTS studies, not least because they allow analysts to test for differences in VTTS measures across modes. However, great care needs to be taken to ensure that all presented modes are within scope for a given respondent, and special precautions will also be necessary to reduce the effects of inertia and the incidence of non-trading.

3.4.6 Small time savings

There is considerable scepticism as to whether SP is an adequate tool for retrieving reliable valuations for STS, with many studies failing to retrieve meaningful effects. At the same time, there also exist arguments as to why such valuations may indeed be very small. With the risk of threshold formation in SP surveys potentially being a contributing factor, one key recommendation in this context would be to use separate surveys for obtaining valuations for STS.

3.4.7 Reliability

For an authoritative statement of methodological challenges and possibilities in the context of reliability, see the recent Dutch scoping study (de Jong et al., 2007) and the DfT-commissioned review study by ITS, Imperial College & Bates (2008).

3.5 Conclusions

To some extent, the AHCG study produced unsatisfactory results in relation to a number of topics that are of particular importance, notably the asymmetry between gains and losses and the valuation of STS. Given these problems, and the concerns, at least in some corners, about the underlying design and survey context, the question needs to be raised as to whether it would be wise for any new study to rely exclusively on the exact same methods.

This is especially true given recent methodological developments. which may be able to produce more reliable results while allowing some scope for cost reductions.

3.5.1 Survey work

In terms of data, the current position is informed particularly by SP data, and to a minor extent by RP data, from the VTTS studies of the 1980s (MVA, ITS & TSU, 1987) and 1990s (AHCG). Looking back upon these studies, from the perspective of 2009 best practice, an immediate reaction is that little has perhaps changed, certainly when it comes to fundamental aspects of method. Irrespective of whether Option A or some form of Option B is in due course commissioned, it is almost inconceivable that this will not be heavily reliant upon SP surveys of travellers faced with trade-offs between journey time and journey cost. In short, the essence of the experimental approach adopted in previous national VTTS studies remains, in our view, perfectly valid in 2009.

Of course, the intervening years have witnessed considerable research effort, and significant methodological developments. In our view, these developments relate to relatively detailed aspects of practice, and do not imply a substantive challenge to the fundamentals. That is not to say that these developments are not significant and should not be taken account of. Rather these developments might be seen more as issuing a challenge to analysts to uphold the empirical rigour and robustness of application, rather than the methods per se. In this vein, we would highlight the following as areas where developments have taken place, and 1999 methods would perhaps fall short of 2010 best practice.

Experimental design

The AHCG employed a very specific non-orthogonal design plan, referred to as the „Bradley' design, which was also used in the Netherlands, Denmark, Norway and Sweden. Design methods have shown considerable progress since 1999, and there are now a range of options at the disposal of the analyst, allowing not only greater customisation, but permitting analysts to obtain more robust findings from smaller datasets, leading to significant cost reductions. This should not be taken to imply that the Bradley design has been deposed; quite the contrary, this design has been applied to the ongoing Norwegian VTTS study. Nevertheless, shortcomings of the experimental design have also been highlighted. In scoping an approach to Phase 2, the challenge that faces us is that different design methods have their proponents and opponents, practice has not clearly unified behind one particular design method, and different methods may give rise to different estimates of VTTS.

Experimental format

The recommendations of AHCG were based largely on an „abstract' binary choice experiment trading time against cost. Whilst there is no apparent consensus on a single best experimental format for VTTS, there is evidence that valuations may vary according to a range of design features including the choice context (i.e. abstract, or mode/route/destination, etc.), the number of alternatives, whether the choice is couched in terms of a „reference' alternative (usually the current choice of the traveller), and the realism/relevance of the experiment to the traveller's actual experience. The challenge for our scoping exercise is that these influences on VTTS may be context specific, and the nature and extent of variation in VTTS difficult to predict.

Survey implementation

This area is less ambiguous. AHCG implemented their SP experiment in terms of a pen and paper questionnaire with postal return. This is rudimentary by modern standards, which typically involve CAPI or CATI surveys. The latter methods enable greater customisation of the experiment to an individual's experience, and can improve response rates and the quality of data. A further issue of implementation is the representativeness of the dataset; this was identified as a weakness of AHCG. Methods of sampling have become considerably more sophisticated, and it is now easier to ensure the necessary coverage of specified segments.

3.5.2 Estimation work

In terms of estimation, the current UK position on VTTS is supported essentially by the binary logit models reported in AHCG (1999). We do however acknowledge that a number of innovations reported by AHCG were well ahead of their time, for example explorations of random parameters and jack-knifing. The state of modelling practice has evolved significantly since 1999, and various enhancements are now at the analyst's disposal. However, in a similar manner to our earlier commentary on data, one might observe that the fundamental elements of estimation method remain largely untouched. That is to say, in 2009, best practice would still draw on models from the logit family, the economic theory supporting the specification of these models for VTTS remains largely unchallenged, and the basic inferences drawn from the model for purposes of appraising VTTS are the same as those derived in 1999/2003. That said, we would highlight the following as areas where developments have taken place, and 1999 methods would perhaps fall short of 2010 best practice, either in terms of their robustness, or the level of detail explained by the model.

Pre-analysis processing

In response to phenomena variously described as non-traders, inconsistent and/or outliers, analysts adopt various practices, including straightforward removal from the dataset or some form of management. AHCG found significant evidence of non-trading, particularly in the case of their toll road experiment, wherein non-traders accounted for up to 30% of respondents. By contrast, the recent Swedish study - using similar methods - eliminated most non-trading by extending the range of bids. Moreover, a challenge in scoping Phase 2 is that methods are not agreed, but different approaches to pre-analysis processing may lead to different estimates of VTTS.

Taste heterogeneity

AHCG was notable in making some initial steps towards the estimation of mixed logit and latent class models, and ITS/Bates reported some of the formative applications of mixed logit models in UK VTTS studies. In the subsequent years, these estimation methods have completed most of the journey from blue skies to leading edge of practice, particularly in the case of the latter, which is the subject of a dedicated WebTAG Unit (3.11.5). Reinforcing the notion that such methods are no longer the preserve of academic studies, we note that the recent Danish VTTS study made extensive use of advanced models of taste heterogeneity. Even if the interest is primarily or solely on mean values, accommodating heterogeneity around this mean may in fact lead to a more reliable mean value. Whilst there would appear to be general agreement that the modelling of taste heterogeneity is informative, there is less agreement on detailed model specification. Since different model specifications may yield different estimates of VTTS, this is a further challenge for our scoping exercise.

Non-linear effects

AHCG devoted careful analysis to the prevalence of non-linear effects. However, their claims regarding the presence of „sign' and „size' effects were challenged by ITS/Bates, and official guidance has retained the status quo of valuing all time savings at the same rate. ITS/Bates did however find evidence of „inertia' and „tapering' effects, recommending equation (3.1) above as a model of VTTS varying by income and journey length (via journey cost) . Whilst similar methods have been applied, and similar phenomena observed, in other VTTS studies - notably the Swiss VTTS study - this is an area of method which in our view is important but unresolved.

4 REVIEW OF EVIDENCE: CAR AND PUBLIC TRANSPORT PASSENGERS

4.1 Introduction

4.1.1 Objectives and scope

The objective of this chapter is to summarise our current understanding of the UK evidence base regarding the value of time for car and public transport passengers, with an emphasis on evidence that has emerged since the second national study conducted in 1994 (HCG et al., 1999) and with a focus on in-vehicle time. The Department has identified a number of specific „areas of concern', to which we have added some others that we feel merit coverage here.

4.1.2 Method

In what follows, a threefold approach is used:

- Update of ITS' previous value of time meta-analyses, including that undertaken for ITS/Bates (2003).
- Interrogation of the meta-data set to address specific issues in a controlled manner at the study level but nonetheless providing quantitative evidence pooled across studies.
- The traditional literature review approach which examines the findings of specific studies.

As outlined in Chapter 2, the early British evidence base and official recommendations were dominated by a few pioneering studies, among which was the first national study (Beesley, 1965; Quarmby 1967; Daly & Zachary, 1975, MVA et al., 1987; Fowkes, Nash & Marks, 1986). Note, however, that the evidence base here is not restricted to national value of time studies, nor indeed to what might be classified as „pioneering' or even „major' studies, but rather we exploit the very considerable amount of high quality behavioural research that has, for many years, been regularly conducted in the UK. Thus we are not primarily interested in national value of time studies, unlike our earlier review of European evidence and practice, but rather in any British empirical study that yields a value of time, even if estimating the value of time itself was not the primary purpose of the study.

4.1.3 Updated meta-analysis

The aim of meta-analysis is to quantify variations in relevant parameters, here the value of time, across different contexts. It is critically dependent upon assembling large datasets of past empirical evidence so as to be able to quantify the main causal factors underpinning value of time variation with a sufficient degree of confidence.

The principal advantage of meta-analysis lies in exploiting evidence from a wide range of studies, yielding insights that cannot be obtained from studies in isolation. Its potential contributions to national value of time studies are as follows:

- It is generally preferable to base recommended values on, or challenge established conventions with, the results of numerous studies rather than a few or a single one.
- It is possible to draw conclusions relating to spatial and particularly temporal variations in valuations that are beyond the scope of a single study; a particularly important one is how the value of time varies over time.
- Results not otherwise in the public domain for reasons of commercial confidentiality can be exploited because the means of analysis maintains their anonymity. Other non-published evidenced can be sourced. It is wasteful not to exploit existing evidence to the maximum extent possible.
- Traditional literature reviews focus on mean values rather than the variation and there is always a risk that a comparison of means, rather than some more systematic quantification of all effects, is distorted by confounding effects.

Of course, meta-analysis cannot examine the value of time at the level of detail that specific discrete choice studies can. There might also be instances of confounding effects, such as longer distance studies offering larger time savings or containing a different blend of journey purposes or income levels. Such possibilities need to be borne in mind in interpreting its results.

The maximum level of segmentation in our meta-analysis is by distance, mode, journey purpose and data type. Thus we collect a value of time as an independent observation where it is either the only value reported in a study or where there are separate values by the above factors for which the meta-model will contain separate independent variables in order to discern their effect.

Our meta-analysis is based upon 226 studies covering the period from the early 1960s to 2008, and these yield 933 values of in-vehicle time and 816 valuations of other variables. It does not therefore focus exclusively on the more recent evidence, since to do so would remove valuable information.

We point out that over time a number of repositories of behavioural information have emerged, which cover but are not exclusive to the value of time. These include the Demand for Public Transport „Black Book' (TRRL, 1980) and its 2004 „White Book' update (TRL et al., 2004), Transport for London's extensive amount of research into the valuation of a wide range of attributes, some of which is contained in its Business Case Development Manual, and the railway industry's Passenger Demand Forecasting Handbook (ATOC, 2009). Our meta-data set covers relevant studies contained in these documents. However, the treatment of some factors, the practical aspects of appraisal and the links with forecasting can and do differ from WebTAG guidance; such issues are not covered in this document.

4.1.4 Within-study interrogation of meta-data set

We can supplement the results of the meta-analysis, which is based on comparison of variations across studies, with evidence of variations in values that occur within studies. Thus for example, we can compare how values of time obtained from a single SP exercise vary by purpose, mode, distance or data type, and assess such evidence that exists across studies. This method controls for extraneous influences whilst also making it quite explicit the amount of data upon which any inferences are based on and allowing more recent data to be isolated. The meta-data set provides a valuable source of information since it is based on what we believe to be a comprehensive trawl of the UK literature, both published evidence and the large amount of „grey literature'. As such, it makes more sense to interrogate the data set to obtain quantification of various effects rather than conduct a blow-by-blow account of even the more significant studies. Given the focus of this review, as opposed to the aims of the meta-analysis more broadly, the discussion that follows focuses entirely on post 1994 evidence relating only to the value of in-vehicle time. This covers 95 studies and 473 values.

4.1.5 Standard literature review

Meta-analysis is focussed on explaining variations in „central values' across studies. There are a range of issues, such as, for example, STS, group size and income effects, which meta-analysis does not tend to cover, not least because in many cases there would not be sufficient observations to support meaningful analysis and because any observations collected would not fit easily alongside the other data assembled. Hence we are reliant on reviewing specific studies which have addressed these issues in order to understand the current state of evidence. Moreover, it would also seem appropriate when considering key issues to focus on the principal studies rather the meta-analysis approach that places equal weight on all evidence.

4.2 Update of meta-analysis

As part of our 2003 study for the Department for Transport on the value of time, we updated meta-analyses that had been conducted in 1997 and again in 2000. The emphasis of all these meta-analyses has been on UK evidence, although the set of variables of interest and obviously the time periods has varied across studies. The emphasis of the 2003 meta-analysis was upon the valuations of in-vehicle time, walk time, wait time and headway. The current update, funded as a small piece of research, covers more recent material and also revisits a range of other attributes, including departure time shift, search time, time spent in congested and free flow traffic, and late arrival time, that were covered in our 2000 study.

This most recent update covers 226 studies. The number of value of in-vehicle time observations increased by 34% from 698 to 933 whilst the proportionate increases for walk time, wait time and headway were 18%, 26% and 32% respectively to 270, 77 and 210 observations. Including the other valuations relating to late time, departure time shift, car parking search time, congested travel time, the model contains 1749 observations.

The model developed to explain how valuations vary across studies is reported in Annex E. We here report a number of key findings.

- The GDP elasticity across all valuations is estimated to be 0.90 (±7.9%), and there is little variation by journey purpose, mode or whether the journey was inter-urban or not. In the previous study the GDP elasticity was estimated to be 0.723 (±43.5%).

- When the GDP elasticity is constrained to be one for business trips, in line with official recommendations, the GDP elasticity for non-business trips is 0.92 (±8.1%).

- The distance elasticity on the value of in-vehicle time is 0.156 (±28.2%). This increases to 0.203 (±29.4%) for car time but falls to 0.115 (±48.2%) for valuations of attributes other than in-vehicle time and late arrival time. The previous meta-analysis obtained a distance elasticity of 0.184 (±31.7%) with a similar increment of 0.075 for car travel and increments of -0.073 for walk and wait values and -0.197 for headway values.

- The current study did not discern a positive effect on the value of time as a result of a journey being inter-urban, over and above any distance effect, but it did find that rail users making inter-urban trips have valuations 27.3% larger than rail users making suburban trips, all other things equal.

- Relative to a base of leisure/off-peak travellers, the category of commuters/peak travellers have values around 11% higher whilst business travel valuations were 110% higher. These findings are consistent with the previous meta-analysis.

- Other notable findings were that RP-based valuations of IVT, walk and wait were respectively 15%, 31% and 31% higher than their SP counterparts, which could be important in planning for future studies, rail time is valued 15% less than car time and this difference will increase with distance but, at least for car users, bus time is valued 18% more highly than car time. Bus users have 34% lower values than car users yet train users have values 43% higher than car users. Car users have a stronger dislike of out-of-vehicle time; compared to all other users, they value walk and wait time 17% higher and service headway 21% higher.

- Compared to SP exercises based on self-completion, CAPI or card presentations, the valuations from adaptive SP exercises, web based surveys and phone based surveys were 33%, 84% and 42% larger. This result could however be something of an anomaly, arising from a shortage of evidence.

- No significant difference was found between car free flow time and in-vehicle time in general whereas time spent in congested traffic conditions was found to be 37% higher. As regards the latter, an argument made in the first UK national study was that congestion could be a proxy for reliability, although our meta analysis found that the congestion multiplier was unaffected by whether reliability was included as an additional attribute alongside.

- The studies also provide a number of insights relating to the valuation of other attributes. The most interesting here are those relating to walk time and wait time, since these also enter official recommendations,

- The value of walk time depends upon the amount of walk time, with an elasticity of 0.097. Similarly, the value of wait time has an elasticity with respect to the amount of wait time of 0.072. Given these elasticities, and also that the distance elasticity varies between in-vehicle time and walk and wait time, the time valuation of walk and wait time will depend on the level of walk and wait time and on distance.

- The following two tables (Tables 4.1 and 4.2) report the time multiplier for walk and wait time across a range of distances and levels of walk and wait time. What can be observed is that the data does not seem to support the premium valuations of walk and wait time that we are accustomed to. Indeed, when the RP multiplier is removed, the values very much appear on the low side; for example, the rail valuation at 200 miles and for 30 minutes would fall from 1.71 to 1.31 whilst the corresponding effect for wait time would be a reduction from 1.57 to 1.20.

Table 4.1: Journey time multiplier for walk time, from meta-analysis

Walk Time	**Car**			**Rail**		
	10 miles	50 miles	200miles	10 miles	50 miles	200miles
2	1.34	1.16	1.03	1.49	1.39	1.31
5	1.46	1.27	1.12	1.62	1.52	1.44
10	1.56	1.36	1.20	1.74	1.63	1.54
20	1.67	1.45	1.28	1.86	1.74	1.64
30	1.74	1.51	1.34	1.93	1.81	1.71

Table 4.2: Journey time multiplier for wait time, from meta-analysis

Wait Time	**Car**			**Rail**		
	10 miles	50 miles	200miles	10 miles	50 miles	200miles
2	1.32	1.14	1.01	1.46	1.37	1.29
5	1.41	1.22	1.08	1.56	1.46	1.38
10	1.48	1.28	1.14	1.64	1.54	1.45
20	1.55	1.35	1.19	1.73	1.62	1.53
30	1.60	1.39	1.23	1.78	1.66	1.57

Finally, arising from application of the meta model in a „forecasting' context, we present implied values of time, in 2008 quarter 4 prices and incomes, across the key dimensions of the meta-analysis (Table 4.3). The absolute values seem quite plausible, the distance effect is apparent and seems quite sensible, particularly relative to a lot of other empirical evidence, and the mode effects also seem reasonable. The relativities indicate there is not a great deal of difference by distance, which is in line with official recommendations in many countries, and that it is the bus market which, not surprisingly, is out of line with other markets.

Table 4.3: Meta-model implied monetary values of IVT (pence/minute in Q4 2008 prices)

	Miles	ABSOLUTE VALUES				RELATIVE TO CAR USERS' VALUE OF CAR IVT		
User		BUS	RAIL	CAR	CAR	BUS	RAIL	CAR
Valued		BUS	RAIL	BUS	CAR	BUS	RAIL	BUS
Commute	2	4.9	9.1	8.8	7.7	0.64	1.18	1.14
	10	6.4	11.7	11.3	10.7	0.60	1.09	1.06
	50	n/a	19.2	n/a	14.8	n/a	1.30	n/a
	100	n/a	21.4	n/a	17.0	n/a	1.26	n/a
Other	2	4.5	8.2	7.9	6.9	0.64	1.18	1.14
	10	5.7	10.6	10.2	9.6	0.60	1.09	1.06
	50	n/a	17.3	n/a	13.3	n/a	1.30	n/a
	100	n/a	19.3	n/a	15.4	n/a	1.26	n/a
	200	n/a	21.5	n/a	17.7	n/a	1.21	n/a

Notes: values are for outside the South East, RP data, 9 comparisons and other terms at their base level; n/a denotes this distance is not applicable for the mode in question

In summary, we conclude that:

- **The meta-analysis has yielded a plausible and precise GDP elasticity.**
- **The distance elasticities also seem reasonable.**
- **A number of modal and purpose effects are apparent, broadly in line with expectations and with the findings of the 2003 meta-analysis.**
- **However, the results for walk and wait time are a little surprising. They suggest that the values of walk and wait time are somewhat less than we are accustomed too, particularly if based solely on SP evidence. For this reason, and also for other reasons discussed below, we feel that there is a case for commissioning further research on the valuation of walk and wait time.**

4.3 Specific evidence and 'areas of concern'

In this section we primarily consider evidence relating to the „areas of concern', whilst broadening our scope to include two other dimensions of interest, namely „types' of time (congested vs. free flow, walk and wait) and RP vs. SP methods.

In addressing these issues, we have interrogated the meta-data set and reviewed specific literature. For each area we conclude whether since the 1999 study:

- there is no new evidence, recognising whether there was any existing evidence or not and thus whether any gaps in understanding persist;
- there has been new evidence but there already existed a significant body of evidence;
- there is new evidence and this challenges existing evidence or fills a gap in understanding.

We also comment on what we believe to be the research that is warranted on the basis of current understanding.

4.3.1 Income segmentation

The first British national value of time study (MVA et al., 1987; p122) concluded that, *"...... we have clearly demonstrated the existence of an income relationship, which has never been done before with any conviction" and that "the value of time as a proportion of income is a decreasing function of income, rather than a constant as has hitherto been assumed"*.

Many subsequent studies, both in Britain and elsewhere, have recovered income elasticities which fall short of one by some margin. For example, ITS/Bates (2003) also estimated the income elasticity alongside the duration elasticity. The income elasticity was 0.36 for commuting and 0.16 for other, the latter in particular being rather low.

Subsequent work, not only in the area of value of time but also the valuation of environmental factors, follows the same pattern of recovering income elasticities well below one, with 0.5 being a representative figure. The M6T study (Wardman et al., 2008) recovered a monotonically increasing value of time across nine household income groups, as reproduced in Table 4.4. However, despite this being an impressive pattern of results, the implied income elasticity is less than 0.5.

As part of the 2003 work, we reviewed the cross sectional variation in the value of time with income (Wardman, 2003). What was apparent, although admittedly from a single study (Fowkes, 1986), was that the per-person disposable income elasticity was much larger, and around unity, when the income segmentation was done using individual rather than gross household income. We would expect the specification of income to affect the results. Relevant factors are whether income is household or individual, gross or net of tax and the number of household members (as well as size and composition of the travelling group; see section 4.3.2). As far as we are aware, very little attention has been paid to the most appropriate measure of income to use in examining cross-sectional variations in the value of time. Wardman (2001) concluded that, *"There are also doubts about the cross-sectional income elasticities obtained, and whether the income elastictity over time is as low as the cross-sectional evidence implies. In part this relates to whether the use of household income has distorted the results"*. That study also reports joint estimation of the Tyne Crossing study conducted in 1987 and in 1999. The evidence indicates that the inter-temporal income effect is low, given that the value of time exhibits little or no growth over time. This contrasts with the results of the meta-analysis where an income elasticity of one is obtained.

More recently, Game 1 and Game 3 of the HCG study were repeated as part of the M6 Toll Study (Wardman et al., 2008), which concluded *"…….. that there is no strong support for differential rates of change in preferences towards toll costs and other motoring costs between 1994 and 2006, but that there is support...for a diminishing value of car travel time for motorists over time"*.

Table 4.4: VTTS across income groups, from M6T study

	Median	Mean
<£10k	7.55	6.82
£10-29k	8.32	7.25
£30-39k	9.38	7.90
£40-49k	9.39	8.08
£50-59k	9.60	8.54
£60-69k	9.85	9.09
£70-89k	10.19	9.17
£90-99k	11.59	10.48
£100k+	13.10	11.76
Not Known	9.33	8.07

Note: Income expressed in terms of gross household income

In summary, we conclude that:

- **Fresh evidence continually emerges to add to the existing evidence on cross-sectional variations in the value of time with gross household income. This evidence consistently indicates an income elasticity less than unity.**
- **The most appropriate income measure to use has not been examined in the detail that it deserves and should be addressed in further work.**

- **The repeat studies, of which there have been two in Britain, are in stark conflict with the meta-analysis since they do not support the value of car travel time increasing over time.**

4.3.2 Group size/composition

On group size and composition, a number of issues are relevant:

- The marginal utility of time might vary between solus and group travel, either for altruistic reasons or because travelling accompanied incurs different utility, whilst there are also cost implications in terms of shared costs.
- Group decision making processes may need to be considered.
- Passengers do not necessarily have the same value of time as drivers.

In the extensive segmentation process undertaken in the first study, car occupancy featured as an incremental effect on time and cost coefficients. However, no clear pattern of results emerged and there was some uncertainty as to what the valuations represented.

AHCG state that, *"In the 1980 study the car passenger valuation was implicitly estimated at close to zero as it was a car driver VOT which was ascertained, and occupancy had little effect"*. However, if the car driver has been asked to evaluate time and cost trade-offs and implicitly to value their own time, then it cannot be inferred that the car passenger valuation is zero, just that it has not been valued. Whilst we might expect the value to be influenced by whether the traveller is accompanied, we would have no idea without asking about the extent to which the respondent has included the time saving benefits to other occupants in their decision making. The AHCG study segmented by occupancy but found different effects on commuting (positive) and other (negative). With regard to passenger values of time, AHCG stated that, *"To our knowledge, there had been no concerted effort made to research this before"* (although we suggest that this apparent lack of „effort' might have reflected an assumption that a market for passenger values of time did not exist!). Indeed, AHCG would seem to have dismissed such an assumption - if it did indeed exist - finding car passengers to have significantly higher values than for drivers for business travel, but marginally lower values for other purposes.

Our impression is that empirical studies pay little attention to the issue of occupancy and group travel. Unlike the study reported in the section on trip duration, which distinguished the cost coefficients between solus and group travel, the general position is a rather cavalier attitude to the specification of cost and group size. Similarly, the marginal utility of time could well be somewhat lower if travel is accompanied, independent of any cost related issues, but it is possible to envisage situations, such as when looking after young children, that this might not be the case. However, this issue is not routinely examined in studies yielding value of time estimates. The position regarding accompanied travel seems to be that there is considerable uncertainty as to what a respondent's implied valuation represents. The AHCG study stated that *"it is evident that the drivers are not taking full account of their passengers' values when making their choices"*. The problem is we do not know to what extent they are being considered. What is needed is a clear specification of what the respondent has to take into account or a more detailed probing of what they have actually taken into account.

An additional issue is that the outcome of group decision does not necessarily have to reflect the average valuation of the group; the group decision might be dominated by a high value of time member behaving selfishly whilst the mean valuation may be much lower. Generally, very little is known about the dynamics of group decision making. However, if the valuation that is of relevance is the WTP of each person on a particular mode, the process for choosing their mode is only relevant insofar as changes in time affect demand. If we believe we have a good idea of the elasticities to time and cost, we can concentrate on estimating how the value of time differs with group size.

In summary:

- **There needs to be greater clarity in what the respondent is valuing for whom. This might require debriefing questions to identify to what extent an individual considered the benefits to others, although debriefing questions might usefully cover a range of attitudes to group travel and decision making. In contrast, studies tend to be clearer on who is paying.**

- **Once this clarity regarding where the time saving benefits accrue and where the costs are incurred is established, it is a straightforward matter to segment the marginal utilities of time and cost appropriately.**

- **The explanation of group decision making might well be complex. However, a clearly specified SP exercise could isolate such processes to yield a valuation that represents the welfare gain that would accrue to a time saving. This would seem to argue for an SP exercise that was not based around actual choice contexts, where group decision making prevailed, but on abstract choices choices involved unlabelled alternatives.**

4.3.3 Trip duration/distance

One of the most consistent effects apparent in the literature is that the value of time increases with duration. The meta-analysis has recovered what seem to be plausible elasticities of the value of time with respect to distance, essentially by comparing valuations across studies with different mean journey lengths. The elasticity obtained is broadly in line with what has been obtained in our previous meta-analysis, and hence there is an additional element of consistency here. However, we note several points:

- The use of mean journey length may have impacted on the results
- The results do not indicate to what extent the variation in the value of time by „distance' are driven by variations in the marginal utility of time by time and variations in the marginal utility of money by cost
- Distance does not correlate perfectly with either time or cost, and is not the driver of variations in the marginal utilities of time and money

- There could be confounding effects, such as larger time savings or different purpose or income mixes for longer distance journeys.

We therefore need to examine specific studies that have segmented by time and/or cost band or that have specified non-linear functions which effectively allow such variation.

We first recount the evidence of the Bates & Whelan (2001) re-analysis of the 1999 AHCG data. The directly estimated „distance' elasticity, estimated in this instance to journey cost, was 0.42 (±31%) for commuting and 0.31 (±17%) for other trips. Allowing the value of time to vary with the level of cost provided a better fit than allowing it to vary with the level of time. However, applying these elasticities over distances ranging from 2 to 200 miles would imply very significant elasticity variation.

Table 4.5: VTTS for inter-urban motorists' route choice, from M6T study

	VTTS	Elasticity
30m	7.54	0.06
60m	7.93	0.09
120m	8.57	0.14
180m	9.12	0.17
240m	9.62	0.20
300m	10.08	0.22

Table 4.6: VTTS for inter-urban motorists' route choice by journey purpose, from M6T study

	Business		Commuting		Other	
	VTTS	Elas	VTTS	Elas	VTTS	Elas
30m	6.1	0.19	6.4	0.54	5.9	0.12
60m	7.1	0.25	10.7	0.92	6.5	0.15
120m	8.6	0.31	22.6	1.23	7.3	0.17
180m	9.8	0.35			7.8	0.19
240m	10.9	0.37			8.3	0.20
300m	11.9	0.39			8.7	0.21

In contrast to this, Wardman et al. (2008) in the context of inter-urban motorists' route choice found that the sensitivity to toll charge and fuel cost, specified for a corridor covering the M6T, hardly varied across six actual journey time bands but that a monotonic, increasing effect was apparent when the time coefficient was segmented according to the same duration bands. The values of time for different journey times and the associated elasticity implied by the SP model were as given in Table 4.5.

Subsequent work, on a slightly smaller data set, found different distance elasticities by journey purpose, as shown in Table 4.6.

Wardman et al. (1997) analysed 1080 RP choices between rail and car for inter-urban journeys. The form of utility function estimated was of the form:

$$U_i = \sum_m \beta_{im} X_{im}^{\lambda_{im}}$$

whence $dU_i = \sum_m \beta_{im} \lambda_{im} X_{im}^{\lambda_{im}-1} dX_{im}$ so that this study allows the marginal utility of time to depend upon the level of time and the marginal utility of money to depend upon the level of money. The data was used to estimate β and λ (a zero value of λ implies a logarithmic function $U_i = \sum_m \beta_{im} \ln(X_{im})$. Time and cost were specified for the end-to-end journey. Separate terms were specified for both train and car and, for cost per person, for solus and group travel. The results obtained were as given in Table 4.7.

Table 4.7: Time and cost coefficients for solus vs. group, from meta-analysis

	β Coeff (t)	Power		Coeff (t)	Power
Cost-Car Alone	-0.5120 (7.3)	λ = 0.3	Time-Car	-2.7621 (8.9)	Log
Car-Car Group	-0.0828 (3.3)	λ = 0.5	Time-Train	-1.6512 (6.6)	Log
Cost-Train Alone	-0.0448 (7.2)	λ = 0.6			
Cost-Train Group	-0.0012 (4.2)	λ = 1.1			

Whilst Wardman et al. offered no specific insights on how the value of time varies with „distance', the indication would seem to be that both marginal utilities should be allowed to vary.

Daly (2009) states that, *"A recurring feature of recent travel demand modelling studies is the focus on variation of the sensitivity of travellers to changes in travel cost or travel time. Specifically, it has consistently been found that sensitivity appears to decline as trip length increases".* Indeed, it is often found that log of cost formulations perform significantly better than linear terms, particular in behavioural studies. It is concluded that: *"Evidence from value-of-time studies and from large-scale forecasting studies is that the value of time increases with trip length, which would imply that cost damping needs to be applied more strongly, or only, to the monetary cost component".*

Since a number of studies in our present meta-data report values segmented by distance band, this allows for a very controlled analysis of the impact of distance on the value of time. We can estimate the implied distance elasticity (β) from a regression model of the form:

$$\ln\frac{VoT_R}{VoT_B} = \alpha + \beta \ln\frac{D_R}{D_B}$$

R denotes a „reference' value of time and B is some „base' value, here taken to be for the lowest distance (D) for which a study provides a separate value. If a study provides values for four distance bands, then three ratios would be specified relative to the base. Data is pooled across all studies which provide values of in-vehicle time that differ only according to distance. Our meta-data set yields 300 such observations and the results obtained are as shown in Table 4.8.

Table 4.8: Values of time by distance, from meta-analysis

Constant α	-0.001 (0.0)
Distance β	0.196 (7.4)
Adj R^2	0.20
Obs	300

However, several issues here need to be borne in mind with regard to distance effects. The choice models used to estimate the value of time will, if a linear-additive utility function is adopted, force the elasticities to increase with distance since the elasticity to some variable X is given by $\eta_X = \alpha X(1-P)$. To the extent that elasticities do not vary much across distance bands, or not to the extent forced by the linear-additive function, there would need to be some modifying effect which, for example, a logarithmic function would introduce, since then the elasticity is $\eta_X = \alpha(1-P)$. However, this would force the marginal utility of time to fall with journey time, yet there are very sound reasons why the marginal utility of time might increase with journey length rather than fall. For example, longer journeys involve more travel time and therefore less time for other things. We might then expect time savings to be relatively highly valued. Similarly, discomfort might only set in at longer journey lengths and can be expected to increase with duration, whilst it can be inferred that the opportunity cost of time spent travelling can be expected to be large if an individual is prepared to invest a large amount of travel time to consume an activity.

There is the related argument of proportionality, whereby a given time saving will have lesser value on a longer distance journeys. This has intuitive appeal. Whilst a logarithmic function implies a particular form of proportionality, there has not been detailed analysis of such effects. Similarly, SP exercises which deal with longer distances tend to offer larger time savings which could impact on the implied distance effect.

To summarise our findings on trip duration/distance:

- **It is clear that the value of time increases with journey distance, and that there are no methodological barriers to the examination of such effects in future work.**
- **However, there is little explicit evidence on whether an individual would value a specific time saving more or less on a longer journey and why. Nor is there direct evidence on individuals' attitudes to proportional time savings.**
- **There would seem to be some uncertainty as to the cause of the variation in the value of time, and confounding effects from income, the size of time savings offered and purpose need to be isolated, whilst proportionality arguments also need to be tested.**

4.3.4 Trip purpose[7]

The meta-analysis provides plausible results for how the value of time varies with journey purpose. The segmentation by commuting, business and other is the most common segmentation in empirical work, although a peak versus off-peak split is used in urban studies (especially with the car mode, where the average occupancy varies by purpose, and the mix of purpose is different by period). Although finer gradations into specific purposes are possible, and were reported in the second national study, they are comparatively rare.

The table below (Table 4.9) reports the findings from our meta-data set for different journey purposes where the only difference between the valuations reported in a specific study were by journey purpose. Since 1994, the data set contains 36 cases where a study yielded values of time for commuting and leisure, with everything else such as mode and distance the same. On average, commuting values of time exceed leisure values of time by 15%. This backs up the meta-analysis results, although the ratio of peak and off-peak values is somewhat larger.

Table 4.9: VTTS by trip purpose, from meta-analysis

		Ratio	SE	Obs
Commute	Leisure	1.15	0.08	36
Peak	Off-Peak	1.40	0.11	16
Commute	Shopping	1.16	0.16	15
Shopping	Leisure	0.84	0.14	13

[7] Note that the scope of this scoping study does not include business.

The data set confirms that there are few segmentations of the leisure trip purpose. It is only for shopping where there are significant numbers of observations. The results indicate commuting time is more highly valued than time spent travelling for shopping purposes, whilst there is some evidence that shopping time is valued slightly lower than leisure time.

Summarising the evidence on trip purpose, we note that:

- **Segmentation by journey purpose is routine and straightforward, and variations between commuting and other purposes exist.**
- **It would be a simple matter to segment by commuting and other in any further work; the research issue is whether finer sub-divisions of the other category would yield useful results.**
- **Whilst it is possible to distinguish journey purpose from income and duration effects, a confounding factor is that commuting journeys often involve time spent in congested traffic conditions which can be expected to have a premium valuation. This effect needs to be isolated from a commuting effect.**
- **As part of any direct, detailed investigation of time relativities, respondents might be asked to directly appraise commuting and leisure travel time, and indeed different forms of leisure travel time.**

4.3.5 Mode

There are two dimensions here; mode used and mode valued. Users of different modes will have different valuations on self-selection grounds; those with higher valuations will choose faster but more expensive modes or will have purchased a car to enable themselves to save time. They may have higher valuations because of higher income and lower marginal utility of money or simply because they have a higher marginal utility of time spent travelling. On the other hand, the valuation of a particular mode might vary, largely because of quality effects.

The meta-analyses found variations by mode used and by mode valued. There were only a few studies where only the users varied and everything else was constant. This showed bus users to have much lower values than car users but the results confirm the higher values for train than car users (Table 4.10).

Table 4.10: VTTS by mode used, from meta-analysis

Mode A	Mode B	VTTS of Mode A/ VTTS of Mode B	Obs	SD	SE
Bus	Car	0.67	13	0.17	0.05
Train	Car	1.36	4	0.21	0.10

There are more observations amongst the post 1994 studies relating to how values vary with the mode itself, all else equal (Table 4.11). Bus has a value of time on average 20% higher than car, whilst a combined public transport mode, where bus will be well represented, is 16% higher. Train has a higher valuation than car but it is not significantly different. On average, train time has a lower valuation than bus although again the difference is not significant.

Table 4.11: VTTS by mode valued, from meta-analysis

Mode A	Mode B	VTTS of Mode A/ VTTS of Mode B	Obs	SD	SE
Bus	Car	1.20	37	0.73	0.12
Train	Car	1.08	31	0.72	0.13
PT	Car	1.16	31	0.79	0.14
Train	Bus	0.88	13	0.35	0.10

In summary:

- **Whilst the value of time is seen to vary by mode user, and the variation is strong, this is proxying for other things and is a necessary approximation when conducting meta-analysis. In a detailed study of value of time, the factors driving this variation by mode should be identified and valued.**
- **Of more significance is variation by mode valued. We would expect variation in the value of time by mode to reflect comfort, convenience and security-related issues. Indeed, it has been claimed that increases in the comfort of car travel may have contributed to the low growth in the value of car travel time apparent in repeat studies. There are no methodological barriers to estimating separate values of time by mode.**
- **There has been little research which has valued time spent walking when walking is in itself the main mode rather than a means of accessing or egressing other modes.**

4.3.6 Small time savings

The argument that STS have zero value, for example because little use can be made of them or indeed they are not actually perceived, is a longstanding one with intuitive appeal. This is a part of a more general point that the unit value of time might not be constant, and indeed conventional economic theory would imply that larger time savings would have lower unit value. There are well rehearsed arguments, such as the „threshold argument' (Fowkes' peer review of AHCG (1999)), why STS should still be valued at the standard rate if, because of thresholds in the use of saved time, they have zero value (ACTRA, 1978). Indeed, in the context of appraisal, there is good reason to adhere to the Common Unit Value (CUV) assumption. That said, there might still be an issue relating to the functional form of the value of travel time savings, whereupon we might wish to estimate values to small changes in time. We here focus on the empirical evidence relating to small time variations independently of the arguments surrounding appraisal issues.

The first national study (MVA et al., 1987, p170) concluded that, *"….. our own exploratory surveys and some evidence from other researchers lend support to the idea that small time savings should have a lower unit value, when considering individuals in the short term"*.

The second national study (HCG et al., 1999) found that time savings of five minutes or less did not have any value. This is an astonishing finding, since 5 minute differences between time attributes is not uncommon in urban SP studies. However, if we define STS to be one or two minutes, or even less, then it would be difficult to draw a clear conclusion from the HCG study.

The ITS/Bates (2003) study, which conducted further analysis of the 1999 data, found the sign effect apparent in the previous work to be the result of model mis-specification. However, it is concluded that *"With regard to the "size effect", there is no doubt that the data strongly indicates that a lower unit utility attaches to small time changes (whether positive or negative)"*. It went on to state that, *"there must be some doubt, however, as to whether Stated Preference is a suitable vehicle for carrying out the investigation of responses to small time changes"*.

We are not aware of recent studies that have specifically addressed the issue of STS, as here defined of being up to two minutes but most certainly including less than one minute. However, we are aware of studies that have offered STS when valuing what might be termed „soft factors' or „secondary variables', which are expected to have relatively low values and hence small time variations must be offered if robust valuations in time units are to be obtained. This issue of STS does not seem to be a particular issue in these studies. It is routine practice to value soft factors in terms of equivalent amounts of time, and it is clearly then sensible to offer only small time variations, otherwise the time variation dominates and the SP exercise provides us with little useful information.

Another instance when small time variations are offered is with respect to walk and wait time, since these are by their very nature often small. However, the time savings tend not to be as small as one minute and, as we have noted in preceding discussion, there are concerns about the reliability of walk and wait values obtained in SP studies.

Time savings can also be small simply by chance. Where the SP design is based on difference between two options, then typically small differences will not be selected, with a minimum difference of 5 minutes of in-vehicle time being typical, although as we have said, differences in walk and wait time can be less. However, in order to estimate alternative specific time valuations, some exercises select different sets of journey times for each alternative. Whilst some studies might still use values defined in 5 minute intervals, whereupon the smallest non-zero difference will be 5 minutes, there is still the chance that in some studies differences are relatively small. This is more likely to occur with SP designs where the journey times offered are pivoted around current levels and particularly where percentage variations on current levels are offered.

In RP data, there is the possibility that some of the time differences are small, and hence could be examined. However, any analysis would have to be based on reported data, but whether these are normally reported to the level of detail required, instead of rounding to, say, 5 minutes, is very questionable. Any data collection in studies with the purpose of addressing time savings would have to address such issues

Whilst questioning the extent to which it clarifies the debate on STS, we summarise the available evidence as follows:

- **Small time variations in SP exercises could be more prevalent than we might first think, but we are not aware of separate analysis of their utility effects. There might be scope in identifying and re-analysing some suitable data sets as part of any further work in this area.**

- **There is some concern about whether SP exercises are appropriate for dealing with small time variations, but this has not been demonstrated through detailed appraisal, such as through the use of focus groups or specific studies which have thoroughly tested whether SP exercises purposely designed to look at small time variations are credible.**

- **Some studies naturally offer small time variations because they are dealing with soft factors which have low values. There are also some real world choice contexts where travellers are routinely dealing with small variations in attributes, such as applies in queuing or its avoidance, detouring to avoid congestion, or waiting a short while for the next service to avoid overcrowding. SP exercises could be built around such contexts.**

- **Indeed, many would contend that the „proof of the pudding' would lie in RP data and some „outside-the-box' thinking should be devoted to identifying relevant contexts offering a continuum of time savings.**

- **We are not aware of in-depth research that has examined travellers' reactions to STS, along with other factors such as proportionate savings and savings on long and short journeys or journeys of different purposes[8].**

---

[8] We have made reference in several places to in-depth, exploratory research that might usefully reveal how different time savings in different contexts are valued in relation to each other without necessarily aiming to estimate their monetary value. However, we must point out that we have not made any great attempt thus far to identify such studies.

4.3.7 Reliability

As part of the recent study by ITS, Imperial College and John Bates Services (2008), a review was conducted of international evidence on the value of reliability for passengers. The following summary of UK evidence draws heavily from that review, together with evidence collated in PDFH. Generally speaking, the evidence base on the value of reliability is relatively small, certainly when compared to the valuation of travel time savings. What evidence does exist is based almost exclusively on SP analysis; this reflects the difficulty of observing actual (i.e. RP) behavior in response to experiences of unreliability. Reliability was covered to a limited extent in AHCG (1999), although the results were not adopted in official guidance. Since 1994, there have been investigations of the value of reliability on most of the major modes, and the following commentary seeks to summarise that most significant contributions.

Road

Before proceeding, it is appropriate to acknowledge that AHCG's work (which was restricted to road) devoted some attention to reliability, defined in terms of the probability of delay. Among their results was the finding that doubling the chance of an unexpected delay was equivalent to 13 minutes extra travel time in the case of commuters, or 20 minutes extra travel time in the case of business travellers. Halving this chance equated to a three minute reduction in travel time for commuters, or five minutes for business travelers, thereby highlighting some asymmetries in valuations.

Copley et al. (2002) carried out a SP survey among single occupancy car commuters in Manchester. Among the attributes presented for each alternative were a set of travel times (in a bar chart) and average travel time. These data were used to estimate the „reliability ratio', defined as the ratio of the value of journey time variability - in terms of standard deviation - to the value of journey time, reporting an estimate of 1.3.

On the basis of the available evidence, TAG Unit 3.5.7 recommends the reliability ratio approach, and a specific ratio of 0.7 for all journey purposes.

Rail

PDFH is helpful in collating several post-1994 studies of the value of rail reliability.

SDG's (1995) SP-based study of „Big Delays' invited rail passengers to trade off 5% and 10% chances of big delays (defined as 30-90 minutes for short distance passengers and up to 80-120 minutes for long distance), against slower but perfectly reliable services. SDG found an increasing linear relationship between the value of delay and the value of scheduled journey times, estimating average delay to be worth 7-10 times scheduled journey time. Re-interpreting their model in terms of relative delay, the lateness multiplier moderated to the range 1.5-4.5.

The body of work by Bates et al. (2000, 2001) presented comprehensive evidence on monetary valuations of delay minutes, and the multiplier of delay to scheduled journey time. As regards the latter, they reported estimates of between 1 and 5 depending on journey length and purpose; these are summarised in the following table (Table 4.12) taken from PDFH 4.1.

Table 4.12: Multipliers for delay, from Bates et al (2000)

TOC and Journey Purpose	Value of Mean Delay (p/min)	Value of Travel Time (p/min)	Ratio (excluding Expected Schedule Delay term)
Great Western work	149.7	27.1	5.5
Great Western non-work	41.8	10.6	3.9
Connex commuter	15.5	8.9	1.7
Connex non-work	12.7	6.6	1.9
Northern Spirit work	23.0	19.3	1.2
Northern Spirit non-work	11.7	9.0	1.3
Northern Spirit airport	112.0	23.0	4.9
Central Trains work	8.3	14.8	0.6
Central Trains non-work	8.2	8.5	1.0
Virgin Trains work	86.3	20.5	4.2
Virgin Trains non-work	10.1	9.8	1.0

Source: Table C4.1, PDFH 4.1.

Note: Values of time calculated using PDFH 1997 recommendations and inflated from Quarter 4, 1994 values to Quarter 3, 1999 values by applying a factor of 1.2.

Bates et al.'s results were reinterpreted by KPMG (2000) for purposes of determining appropriate multipliers for the Schedule 8 performance regime, and these are given in Table 4.13. In response to claims that multipliers were too low for some specific forms of journey, the Office of the Rail Regulator (2000) recommended the multipliers given in Table 4.14 for use in the Schedule 8 performance regime for control period 2 (2001-2006).

Table 4.13: Multipliers for delay, from KPMG (2000)

	Full fare and Season tickets	Restricted use tickets
London commuter service groups	2.74	2.20
Virgin/GNER	4.56	1.11
Great Western/Midland Mainline/Anglia	6.05	4.21
Northern Spirit/ScotRail/Wales & West	1.82	1.70
Central/North West/MerseyRail/Cardiff	0.78	1.58
Airport service groups	6.52	6.52

Source: Table C4.2, PDFH 4.1.

Table 4.14: Multipliers for delay, from ORR (2000)

	Full fare and Season tickets	Restricted use tickets
Airport journeys	6.5	6.5
Long Distance High Speed Services	6.1	4.2
All Other Services	2.5	2.5

Source: Table C4.3, PDFH 4.1.

MVA's (2000) study reported much higher multipliers of between 5.25 and 13.75. They also reported valuations of the reliability ratio in terms of delay (i.e. ratio of the value of standard deviation of delay to the value of mean delay) of between 0.16 and 0.6. However, the authors noted some statistical problems encountered in the course of their analysis, and these results have not been widely adopted.

In a recent SP study for DfT Rail (Batley et al., 2008), ITS conducted a relatively large SP investigation, based on some 3900 travellers, and estimated the following delay multipliers, segmenting by distance and purpose. With regards to the lateness multiplier at destination, ITS estimated multipliers for business and commuting of 1.80-3.22, these increasing to 1.88-5.30 for leisure (Table 4.15). The authors concluded that these results offered broad support for existing PDFH and WebTAG guidance, but suggested that 3 might be more appropriate for short distance trips, with a lower multiplier of 2 perhaps applying to long distance trips. This discrepancy between valuations for short and long distance indicates the possible prevalence of a „relativity' effect in relation to reliability.

Table 4.15: Multipliers for delay, from Batley et al. (2008)

	Business	Commute	Leisure
Short	2.70	3.22	5.30
Long	1.80	2.10	1.88

On the basis of the various evidence detailed above, WebTAG Unit 3.5.7 issues recommendations in relation to both the reliability ratio (in terms of delay, as distinct from IVT) of 1.4, and the delay multiplier of 3.

Bus

There has been relatively little work on bus. The most notable contributions are perhaps those of WS Atkins (1997) and Hollander (2006), both based on SP evidence. The former reported the following monetary valuations of standard deviation in both in-vehicle time and wait time (Table 4.16), from which one can infer reliability ratios (for IVT) of around 1-1.8.

Table 4.16: Money valuations of reliability in IVT and wait time (p/min), from WS Atkins (1997)

Variable	Work	Shop	Other	Peak	Off
IVT	1.4	0.8	1.3	1.4	0.9
SD(IVT)	2.0	0.8	2.3	1.8	1.2
Wait	2.6	1.3	3.1	2.6	1.7
SD(Wait)	2.5	1.0	1.9	3.0	1.2

More recently, Hollander (2006) carried out an SP survey among a relatively small number (244) of bus users in York, eliciting a reliability ratio (again in terms of IVT) of only 0.1. In the light of this limited evidence base, WebTAG guidance on bus reliability refers to the reliability ratio (in terms of delay), citing the same 1.4 value applying to rail.

To summarise:

- **Evidence on the value of reliability is much less voluminous than that applying to the value of time. That said, since reliability is an emerging policy interest in many countries, and there is a degree of consensus on methods, there are opportunities for pooling knowledge and evidence across countries.**

- **Turning specifically to UK evidence, there is a disparity between the comprehensiveness of evidence on different modes. There exists a reasonable evidence base for road and rail, but little evidence for bus; there would be a good case for seeking to strengthen the latter.**

- **In the UK, standard guidance has been issued on the value of reliability for different modes, although different metrics tend to be applied to different modes, making modal comparison difficult. For example, road tends to employ the so-called „reliability ratio', whilst rail instead employs the „lateness multiplier'.**
- **Standard valuations of these metrics have been prescribed by WebTAG and other industry guidance, although these are based on central estimates. In practice, the metrics may exhibit a degree of variability, depending on factors such as journey purpose and distance. More generally, it would be sensible to seek to strengthen the evidence base on the value of reliability by segment.**

4.3.8 Types of time

There are a whole range of factors that influence the marginal utility of time spent travelling and hence serve as modifiers to the value of in-vehicle time. Some are what are termed „covariates'; socio-economic characteristics and trip features that lead to different valuations across individuals and contexts. The first national value of time study conducted extensive analysis of such variations, as did the second study, and indeed such analysis can be readily incorporated in any future study. However, more relevant here is the possibility of different valuations according to the features of time itself, rather than of the individual incurring the time.

Such forms of time are: walking time, waiting time, time spent in various degrees of congested conditions and indeed standing, search time, late arrival time and departure time shift. Indeed, these valuations may also vary according to the precise conditions in which they are incurred, so that for example, walking upstairs has a higher value than walking downstairs or on an escalator.

Our understanding is that the Department is not particularly interested in covering, within any future value of time study, search time, departure time shifts and standing time, and that late arrival time would effectively be covered by analysis of reliability.

Congested versus free flow time

The first value of time study distinguished between time spent in congested and free flow traffic, and was one of the first to make such a distinction. Since then there have been a number of such studies and these are covered in our meta-data set. This indicates that, across 29 observations from 9 studies, the ratio of congested to free flow time averages 1.54. It varies little across business (1.43), commuting (1.59) and leisure (1.54). The second national study did not explicitly include congested time in the SP exercises, although the proportion of time spent in congested traffic for the actual journey was included as a covariate.

Wardman, Ibáñez & Pagan (2009) provide a review of international evidence and find that it tends to corroborate the British evidence of a premium of around 50% attached to congested time. The first study to extend to three types of time was, in our understanding, Hensher (2001) which reported a distinction between three types of car time; namely free flow, slowed down and stop/start. In a piece of work for the railway industry's Passenger Demand Forecasting Council, whose purpose was to estimate the effects of car competition on rail demand, Steer Davies Gleave (2003) used a finer gradation of types of time in their SP exercise, and this was adopted by Wardman, Ibáñez & Pagan (2009) in applying the same SP exercise in the United Kingdom and the United States. The results of these three studies were as shown in Table 4.17.

Table 4.17: Valuations for types of time, from selected UK and US studies

Travel Conditions	UK 2006	US 2008	SDG 2003
Free Flow	1.00	1.00	1.00
Busy	1.05	1.03	1.20
Light Congestion	1.11	1.06	-
Heavy Congestion	1.31	1.20	1.46
Stop Start	1.20	1.38	1.67
Gridlock	1.89	1.79	1.96

We must therefore conclude that the value of car travel time depends strongly upon the conditions in which it is specified.

There must also be some uncertainties about what kind of time is actually being valued in SP exercises when it is not clearly specified what the value is. This is a potentially serious issue given the range that the value of time can take across conditions. However, as noted earlier in this chapter, our meta-analysis did not uncover any significant differences between car free flow time and in-vehicle time in general.

The second national study also conducted an SP exercise where lorry access could be restricted or prohibited, there were one or two additional lanes and there was no shoulder or a wider shoulder, and these were traded-off against travel time. The results were however mixed.

As part of the M6T study, there were SP exercises that examined how the value of time for motorists depended upon other factors impacting upon traffic conditions (Wardman & Ibáñez, 2008). The key findings were that:

- The value of time increased by 0.7% for every one percentage point increase in the proportion of HGVs.
- Wide lanes reduce the value of time by 5% and narrow lanes increase the value of time by 9%.
- Travelling on a concrete surface adds 12.2% to the value of time, contrasting with a 9.0% uplift for a high-level jointed surface.
- The presence of lighting reduces the value of time by 9.8%.

- Compared to 3 lanes, 4 lanes would reduce the value of time by 6.9% but 2 lanes would increase the value of time by 10.5%.

Walk and wait time

These are significant components of using public transport modes, and walking is also incurred in making car trips. The first national study and the 2003 update covered walk and wait time but the second study did not.

There is a considerable amount of evidence on the value of walk and wait time, and indeed headway. The meta-analysis examined this and found that the implied walk and wait time valuations in equivalent units of in-vehicle time were lower than we might expect, particularly so if the RP evidence is ignored.

We might speculate that SP exercises can have difficulties in estimating walk and wait time. For example, it might be difficult to convey why walk time or wait times would vary, for example, where walking speeds and distances do not vary and where individuals arrive at their bus stop or train station a few minutes before the planned departure.

Looking at the meta-data set, the valuations of walk and wait time in equivalent units of in-vehicle time are as given in Table 4.18:

Table 4.18: Valuations of walk and wait time, from meta-analysis

Attribute	Type	Mean	Obs	SD	SE
Walk	RP	1.84	39	0.92	0.15
	SP	1.62	257	0.70	0.04
Wait	RP	2.32	27	0.92	0.18
	SP	1.43	63	0.59	0.07
Headway	RP	0.91	11	0.60	0.18
	SP	0.78	208	0.47	0.03

Note: The sample sizes are not the same as for the money values since some studies provide money (time) values but not time (money) values.

These figures reflect the results that have emerged from the meta-analysis itself, with purely SP based values apparently lower than we might expect and indeed lower than official recommendations.

Summarising our findings on types of time, we note the following:

- **As with STS and duration/proportionality effects, there is little evidence on the relativities of different types of walking and waiting time and of car travel time conditions.**

- **There is strong support for the value of car travel time varying with traffic conditions. Indeed, it is important to distinguish such effects to be clear as to what precisely is being valued when changes in car travel time are being evaluated.**

- **Further work, involving exploratory research addressing time relativities, more sophisticated presentation of different traffic conditions in SP exercises and corroboration of the values of different forms of congested time in actual behavioural responses would be desirable.**

- **There is a view that the premium associated with congested travel time discerns an element of unreliability valuation. This needs to be tested.**

- **There are concerns about the values of walk and wait time obtained from SP exercises. It may be that the treatment of walk and wait has tended to be a „little cavalier' and further inspection of the literature might provide more insights in this respect. For example, we might distinguish studies where walk and wait times might be more realistic than in others.**

- **It would seem that careful SP exercises focussed specifically on walk and wait time valuation are required. Corroboration with suitable RP data would be sensible.**

- **In contrast to in-vehicle time, little is known about how the values of walk and wait time vary with the levels they take, with journey duration, with the size of the variations and according to the conditions in which walk and wait time are experienced.**

4.3.9 RP and SP

The meta-analysis indicated that RP valuations were higher than SP valuations, although the difference for in-vehicle time is slight and, as we have seen, the main concerns surround walk and wait time. We interrogated the data set across all years for in-vehicle time, identifying cases where a study provided RP and SP valuations and all other factors were the same. This yielded 28 observations, with a mean ratio of RP to SP valuation of 1.33 and a standard error of 0.14. Of these 28 values, the minimum difference between RP and SP values was -4.8 pence per minute, and in 9 cases the SP value exceeded the RP value, whereas in 19 cases the RP value exceeded the SP value with a maximum difference of 8.2 pence per minute.

Summarising:

- **RP valuations seem to be larger than SP valuations, on the whole, although the discrepancy is not always large. This pattern is consistent with response bias toward cost in SP exercises.**

- **Whilst the focus of any VTTS update study should certainly be on SP, any possibilities for corroboration with RP should not be overlooked.**

- **Such corroboration might usefully extend beyond „first order' mean values of time to „second order' effects and test whether the variations with respect to duration, small savings and other non-linearities apparent in SP practice are confirmed in actual behaviour.**

- **This corroboration is particularly important for walk and wait time if their values are as low as the wealth of British evidence seems to suggest.**

- **There is recent evidence from the United States and Australia suggesting that SP yields lower valuations than RP.**

4.4 Conclusions

This review of evidence on passenger VTTS arises from 3 strands of analysis, namely an update of ITS' previous value of time meta-analysis, further interrogation of the meta-data set to address specific issues, and traditional literature review. In drawing out our key observations, we refer to the specific issues of interest (which include the Department's „areas of concern').

Income

- We note that evidence from national VTTS studies in the UK and meta-analysis are not entirely in harmony, notably to the extent that the former do not support the value of car travel time increasing over time. This issue should be reviewed by any subsequent national study. Moreover, we suggest that the most appropriate income measure has not been examined in the detail that it deserves, and this should be addressed in further work.

Group

- We believe that there needs to be greater clarity in what the respondent is valuing for whom; this could call for more sophisticated qualitative research alongside SP.

Distance

- It is clear that the value of time increases with journey distance. Since there are no substantive barriers to the analysis of such effects, this should be fall within the scope of any update study.

Mode

- Whilst evidence on VTTS typically varies by mode, this is proxying for other effects; see Bates/ITS (2003) for a detailed discussion of this point. We would not actively promote the scoping of a VTTS update to encompass both car and public transport; we will discuss such matters in more depth in Chapter 6. If however it were decided in due course to survey both car and public transport, then the factors accounting for variations in VTTS across modes should be identified and valued, and care taken to distinguish between "mode used" and "mode valued".

- In our view, there is a dearth of evidence on walking, when walking is the main mode. We would recommend the commissioning of fresh research to cover this knowledge gap.

STS

- Whilst sympathetic to the argument that SP may not be entirely adequate for valuing STS, we do not feel that the case has been proven. We would therefore recommend the commissioning of research to address this specific question. Until such a time that the suggested inadequacies are proven, or better methods of analysis are developed, we would advocate the retention of SP for the analysis of STS,

Reliability

- We feel that there is a disparity between the comprehensiveness of evidence on different modes. There exists a reasonable evidence base road and rail, but little evidence for bus; there would be a strong case for seeking to strengthen the latter.

Types of time

- We feel that this is an area where the evidence base is evolving, and it would therefore be premature to draw strong conclusions. There is certainly a need for further work exploring time relativities, more sophisticated presentation of different traffic conditions in SP exercises, and corroboration of the values of different forms of congested time in actual behavioural responses.

Revealed preference

- Although this chapter has relied almost entirely on SP evidence, this is not to overlook the potential role for RP evidence, if only to provide corroboration of valuations drawn from SP.

5 REVIEW OF EVIDENCE: FREIGHT & LOGISTICS

5.1 Introduction

5.1.1 Objectives and Scope

Following on from Chapter 4, the purpose of the present chapter is to summarise the evidence base regarding the value of time for freight and logistics. In contrast to the evidence base for passenger, there is considerably less evidence available for freight. For this reason, this chapter supplements UK evidence with evidence from The Netherlands, which might be considered to offer a reasonable analogy to UK conditions. That said, as we shall describe below, ORR are building models of rail time related costs and ITEA have revised their method for estimating lorry and van vehicle operating costs (WebTAG Unit 3.5.6). So, whereas for passenger not much has happened to official VTTS thinking since the 2001-03 review, that is not the situation for freight.

5.1.2 Method

In this case, there is not sufficient evidence to undertake meta-analysis, and we shall therefore proceed simply by drawing upon specific studies that we have knowledge of. It is worth remarking that evidence from the freight and logistics sector can be subject to commercial sensitivities, and for this reason it may be the case that we are unaware of some bodies of evidence, or have awareness but limited access.

5.2 Understanding what the published values mean

An important challenge in improving the evidence base for freight is to ensure that values are, or can be, subdivided according to our latest understanding of the utilities associated with time savings, so we do not just have a mass of estimates that we leave users to dip in to - usually on the basis that they must all be interchangeable. Expanding upon this point, consider the following two issues.

ISSUE 1. Is the reaction, both in real life and in surveys, long term, medium term or short term? In the long term, constraints on departure and arrival times can be varied through negotiations with the functions providing goods for transport (e.g. a factory production line) and those receiving the goods (e.g. a supermarket distribution centre). Estimates of this long run value can be obtained by pointing out to respondents that they should consider starting out earlier if offered a journey time increase or later if offered a journey time decrease. This may seem arcane, but this IS a key part of VTTS. In the medium term the penalties for early and late arrivals are fixed, though respondents offered the possibility of adjusting departure time may take it. Such values form the core of all VTTS values in the literature. In the short run, the departure time is fixed and delays will impinge fully on arrival time, with its associated penalties. There are such values in the literature, but they do not represent VTTS per se, and have more to do with lateness or (day to day) reliability.

ISSUE 2. Is the firm, both in real life and in surveys, paying per unit of time for resources used in the movement? If the firm is moving beer on own account using its own lorries and beer kegs, its value of time will be VTTS, i.e. include all components. There will be the benefit of getting goods to destination quickly (both to reduce working capital and to avoid stock outs); there will be the drivers' wages; there will be the costs of sourcing extra kegs if each returns less quickly; and there will be costs associated with the lorry used. When interviewing such firms a good estimate of VTTS should result; otherwise corrections may need to be made. If presenting firms with alternative door to door services, with various journey times and costs, as might be sensible for a mode choice experiment, the valuations obtained will not be VTTS. They would include the capital charge on the money tied up in the load and the kegs, but not drivers' wages or changes to Vehicle Operating Costs (VOC). Since the latter two items are the only two items currently recognised in WebTAG guidance, the importance of this distinction can hardly be overstated. When interviewing hauliers or other intermediaries, values for these latter elements might be obtained, together with those of any penalties/bonuses payable to/from the shipping firm for late or early arrival; we shall offer insights on such matters in section 5.7.3, drawn from Dutch as opposed to UK evidence.

In looking at values found in the UK literature since 1994, it should now be clear that we need to understand how each value was derived. Due to the difficulty of obtaining long run values, since respondents will rarely have time to consider re-planning their supply chain during the interview, let alone negotiate with those upstream and downstream, we must regard our target as obtaining medium term VTTS values covering all the elements (as listed in the beer example above) of (dis-)utility, preferably separated out.

5.3 Our starting point is the 1999 report

We repeat here material from the ITS/Bates 2001-3 review for DfT. AHCG employed two different experiments, one of which was analysed both with and without the exclusion of some respondents, see Table 5.1 below. Broadly speaking, results are available for four segments, those being the combinations of LGV v HGV and Hire & Reward v Own Account. The experiments appear to be consistent with what we labelled as "Medium Term" in section 5.2 above.

The first experiment considered the choice between two untolled roads, having different times and costs, as well as differences in other attributes. Estimated values of time were 45p/min for Hire and Reward and 35p/min for Own Account, and these formed the recommended values of AHCG. All else equal, this is the opposite way round to what we would expect. We suspected that respondents might have considered the cost changes in Experiment 1 unrealistic and so ignored them relatively to the time changes. This is because, in order to generate the trade-offs necessary for the SP experiment to work, the slower roads would have to be shown as cheaper, which is counter-intuitive if extra distance, stop-start traffic or signalised intersections were the reasons for the slowness. Generally, all else equal, longer routes should **cost** more as well as having longer journey times. Fast roads, such as motorways, will go against this, attracting traffic to take a longer route, at higher cost, but with shorter journey times. However, depending on the interview context, this will not always appear realistic to respondents in particular situations.

Table 5.1: The AHCG models for the freight SP experiments

Segments	LGV Hire and Reward	LGV Own Account	HGV Hire and Reward	HGV Own Account
Freight Experiment 1				
Observations	362	425	812	381
Rho-square (C)	0.16	0.19	0.18	0.14
VTTS (p/min)	43.5	35.5	47.1	35.5
Freight Experiment 2				
Observations	381	453	833	401
Rho-square (C)	0.44	0.36	0.42	0.45
VTTS (p/min)	19.3	20.8	19.5	33.3
Freight Experiment 2 (excluding those always rejecting current route)				
Observations	273	337	631	311
Rho-square (C)	0.38	0.24	0.24	0.27
VTTS (p/min)	15.1	17.7	20.5	59.3

Notes:

(i) „Observations' are not the same as „respondents'. There were a total of 270 respondents, and so an average of 7.33 observations per respondent to Experiment 1 and 7.66 to Experiment 2.

(ii) VTTS is in end-1994 prices

The second experiment overcame that problem by charging a toll to use the quicker (current) route as against a slower, free, alternative route. This is believable, but causes a different problem, an anti-toll bias. This appears to be the case in this experiment since the untolled alternative always has a positive Alternative Specific Constant, implying it is preferred over the current route all else equal. The results from Experiment 2, therefore, must be treated with some caution. Except for the HGV Own Account sector, the typical VTTS found is about 20p/min, which was consistent with drivers' wages plus VOC at that time. The HGV Own Account value is 33p/min, with a 95% Confidence Interval of 20p/min to 46p/min. This greater value for Own Account is consistent with the reasoning given in section 5.2. The overall average over the 4 categories used is 22.4p/min. This is consistent with the reported value of

21.1p/min for a similar 1993 Accent/Hague study (see Accent/HCG, 1999) and with our interpretation of current appraisal practice (i.e. wages plus VOC).

However, there are some important features of this data. 25 per cent of the sample refused to trade time for money at any of the rates offered in the SP. The above results depend critically on the plausibility of the responses of this “low time value”/”anti-toll” group. If the non-traders are dropped, the precision of the estimates is reduced, and the HGV Own Account Value rises to 59 pence/min, while the other categories change little. The reported Rho-square statistics suggest that the Experiment 2 models were considerably superior in fit, and the discussion at the bottom of p232 of AHCG implies that Experiment 2 results are to be recommended, supported by Accent/HCG (1999) results. Subsequently, it appears that they decided to only accept results from non-toll experiments, although the arguments given for that decision all relate to car travel.

5.4 Studies between AHCG and ITS/Bates (2003)

5.4.1 Tweddle, Fowkes & Nash (1995)

Tweddle, Fowkes & Nash (1995) report results from a survey of Anglo-Continental freight movement prior to the opening of the Channel Tunnel. The experiment is essentially “Medium Term” in the sense of section 5.2 of this chapter. As usual, values of time were presented as percentage reductions in the freight rate required to compensate for a unit of extra travel time (with a day comprising 9 such units). Additional calculations here give the implied values of time in 1995 prices, per vehicle, as:

Median:	£33/hour	55p/min
First quartile:	£9/hour	15p/min
Third quartile:	£108/hour	180p/min

Since higher valued commodities tend to get transported internationally, the relatively high VTTS above are not unexpected (particularly if VTTS is related to the cargo value). Some element of the reported spread will represent residual variation of commodity type, but much will be due to other sources (e.g. urgency). We take this as strong evidence that there are influences on actual VTTS that are not accounted for by drivers’ wages and VOC.

5.4.2 Fowkes & Whiteing (2003)

Fowkes & Whiteing (2003) report work for the Highways Agency under their Understanding Travel Behaviour programme. Respondents were road freight shippers or Hire and Reward hauliers. Each was told that the cost for one of their current movements would double due to the imposition of a toll, but that a cheaper (also tolled) alternative was available via a slower route, giving a later arrival. As the possibility of starting out earlier was not allowed, we expect values higher than pure VTTS, i.e. short term values in the sense of section 5.2. These were referred to as the Value of Delay Time (VDT). Also estimated was VSH, a schedule delay not involving a longer journey time, but including the penalties for late arrival (which would also be present in VDT). By subtracting VSH from VDT we can hope to get a rough estimate of pure VTTS. The overall value of delay time (VDT), for the 40 interviews, was 107p/min in end-2000 prices. The overall value of the schedule delay time (VSH) was 66p/min in end-2000 prices. The derived VTTS is therefore 107-66 = 41p/min. This correction should give us an approximation to the medium term values that we desire.

The split by commodity suggests that the values may only be as high as they are because of some specialist products involved (some becoming difficult to unload if aboard the lorry for too long). Another relevant point to be borne in mind is that the traffic was predominantly long distance, averaging at 282km. In terms of the estimated values, the distance split is as follows.

Table 5.2: Freight valuations by distance

(p/min)	***Delay time VDT***	***Schedule delay VSH***	***Hence Value of Time VTTS***
DISTANCE <250 km (av. 133 km)	90	59	31
DISTANCE >250 km (av. 362 km)	125	74	51

In the UK in 1995 the average length of haul for a HGV was 88km. Extrapolating the above figures suggests a VTTS of 27p/min for an average UK HGV movement.

The most interesting split, though, was probably by respondent type, where (for the first time we know of) third party traffic (i.e. sent via a haulier) was split according to whether the shipper or haulier was interviewed. The third party shipper evinced a low value of time, presumably ignoring any effects on lorry utilisation, driver effects, or the possibility of knock-on effects on other movements. Hauliers and Own Account operators gave much higher valuations, despite shorter and cheaper journeys being involved.

Table 5.3: Freight valuations by respondent type

	Av. Cost (or Freight Rate)	**Av. Dist**	**Delay time (p/min)**	**Schedule delay (p/min)**	**Value of Time (p/min)**
Own account	£227	237 km	169	126	43
Haulier	£298	287 km	155	87	68
3rd party shipper	£327	321 km	37	31	6

Our interpretation of the above is that the study's average VTTS is being depressed by inclusion of shippers who are not allowing for drivers' wages etc. The Own Account figure is probably the best figure for us to take forward, although adjustment is necessary to take it from a 237 km value to an 88 km value. That interpolation gives a VTTS of 29p/min. This is not actually lower than the overall average (27p/min) but is sufficiently close to offer support for that figure. At some £16-18/hour, this is a little above what would be expected from the wage rate plus VOC, but not sufficient or sufficiently well founded for a recommendation to alter the official method. Consequently, with this UK evidence and reports from other countries, ITS/Bates (2003) recommended keeping the current method.

5.5 UK studies since ITS/Bates (2003)

5.5.1 The BAH/ITS study for the Strategic Rail Authority in 2003/4

In 2003, SRA commissioned a study from a consortium of Booz Allen Hamilton and ITS, Leeds, directed principally at rail freight and potential rail freight. As we will see, each partner was to use their own method of analysis. Unfortunately, the reconciliation of the methods at the end of the project was rushed, and no easily-usable digest of results was produced.

The BAH method involved asking respondents to build up their VTTS element-by-element according to the costs that would result if journey times were increased. Some re-planning was allowed, so it can be accepted as a medium term approach in the sense of section 5.2 of this chapter. However, the many rail users interviewed reported no driver costs. For siding-to-siding movements no lorries would be involved, and for movements involving road collection and distribution (C&D) it would be assumed that the delay would have been in the rail leg. No attempts were made by respondents to quantify any extra costs of delay to the train operator (driver costs, loco costs etc.), as these were invisible to them. They would not expect the Freight Operating Companies (FOC) to charge them more if the train was delayed in transit. Perhaps surprisingly, it proved difficult in eliciting ANY costs from some respondents. For bulk commodities, the capital servicing charge for an hour's delay is very small per tonne. Where firms own the wagons used, it will generally be the case that there will be very great costs resulting from a delay, since extra wagons will need to be hired or bought to maintain the service (wagons being very intensively used). Some of these extra costs could be deduced from replies and added in, but there is some arbitrariness in which costs were actually considered.

ITS again used its Adaptive Stated Preference approach (LASP). Particular provision was made for respondents to react to scenarios presented by rescheduling the departure time earlier, suggesting that the VTTS are medium term. Respondents chose between two door-to-door services, whether by rail or road, and so did not have to consider drivers' wages or VOC in their decisions; to get a road VTTS these costs would (in principle) have to be added. However, it is not clear that the BAH values never included such costs, making it difficult to draw definitive conclusions.

Table 5.4, below, has been prepared by ORR, and seeks to compare the results of the two methods. For Finished Merchandise and Express there is little difference between the VTTS from the two methods, although the BAH method (referred to as „Revealed Cost') was higher than the ITS method (based on SP). This suggests that the BAH method was including drivers' wages etc. for these commodities, whereas the ITS method was not. For all other commodities the ITS values are considerably above those from the BAH method. An internal review of the ITS results suggested that for bulks they were rather high, but that there were real costs somewhere in the system that would need to be included in VTTS and that the BAH method had missed some of these. For rail, this would include costs of greater wagon and loco provision, even if this fell on the freight operating company (though the more normal case is that wagons are leased by the shippers). Some increment to the BAH values is therefore necessary to get the VTTS values we want. Note particularly that the wagons used for petroleum and chemicals movements on rail are very specialised and expensive, making the BAH figure doubtful.

Table 5.4: Results from the BAH/ITS survey for SRA in 2003/4 (in £ per truckload per hour, 2004 prices)

	Coal	**Bulk petroleum and chemicals**	**Other bulk**	**Automotive**	**Finished Merchandise**	**Containers**	**Express**
% of total traffic	**36%**	**7%**	**26%**	**1%**	**3%**	**26%**	**0%**
Freight rates £ per tonne	6.50	5.08	6.63	71.50	28.80	14.20	17.35
Revealed cost (RC)							
Loaded	0.02	0.26	0.13	4.37	16.9	2.55	63.41
Unloaded	0	0	0	0	0	0	0
Based on	Transit inventory	Transit inventory	Transit inventory	Transit inventory	Transit inventory plus potentially small amount for consumption inventory	Transit inventory	Local distribution/ main depots
Stated preference (SP)							
Loaded	2.4	8.4	1	24	12	24	60
SP as % of RC	12000 %	3231%	769%	549%	71%	941%	95%
Combined							
Loaded	1.08	5.6	0.37	11.71	14.04	9.91	61.72
Unloaded	0	0	0	0	0	0	0
	Coal	**Bulk petroleum and chemicals**	**Other bulk**	**Automotive**	**Finished Merchandise**	**Containers**	**Express**

The “combined” values from the two methods are given towards the bottom of the Table 5.4. Although presented “per truckload”, not all have had drivers’ wages and VOC included (since these only arose in the BAH survey, and only for respondents talking about a current road movement).

5.5.2 The 2009 AECOM/ITS study for ORR

The jury is still out on how to interpret the BAH/ITS findings. ORR have been particularly interested in clarifying the position (and so producing Table 5.4 above), and DfT Rail have also now renewed interest. They decided that they should commission more work reviewing the BAH method, and ITS was part of that team (AECOM & ITS, 2010). This review attempted to tease out more costs than those identified by BAH, but proved to be a difficult process. Because ORR wanted to use its own estimates of locomotive and wagon costs, the published VTTS generally only includes the capital cost charge for the delayed goods.

5.6 Reliability

Consideration of reliability involves many more complications that cannot be addressed here. Suffice it to say that the HA, BAH/ITS, and AECOM/ITS studies all have reliability as a central interest. Analysis has suggested that reliability can be well presented in experiments as “spread” (of arrival times) and modelled in relation to “lateness”. Little has been published of the results, and some effort would be required to convert them to a common basis. Again, ORR cut out elements of VTTR from their published figures. In this case, reliability is a short run concept, in the sense of section 5.2 of this chapter, being unpredictable and requiring a short term response and facing fixed penalties and rewards for late and early arrivals.

5.7 International freight

Broadening our perspective beyond the UK, let us now seek to draw some parallels with The Netherlands, which also has assembled a body of evidence on VVTS for freight and logistics.

5.7.1 National freight value of travel time surveys

Prior to 1992, in Dutch CBA the freight value of time was based on transport costs (including staff and vehicles). In 1992, the results of the first national VTTS survey became available, based on SP experiments among shippers and carriers. These values have been used until the values from the update survey in 2004 (again SP-based) were published. In 2009, a new SP-based survey on both the passenger and freight VTTS and the value of travel time reliability (VTTR) started (the survey was designed in 2007, and is also being used at the moment in a passenger and freight survey in Norway). For freight, the main reason for carrying out this new survey is to include travel time reliability. The freight VTTS is included in the SP experiments in an effort to avoid double counting. The transport modes included are road, rail, inland waterways, sea and air transport.

5.7.2 Which freight value of time?

Several freight VTTS studies in Europe arrive at estimates of 30-50 Euro per road transport per hour (de Jong, 2008). These values include both the operating cost component of VTTS (labour, fuel, vehicles) as well as the component related to the cargo itself (such as the capital costs of the inventory in transit). Some other studies, such as the review by Bruzelius (2001), focus on the cargo component, which for most shipments will be quite small, unless the goods have a very high value, deteriorate very quickly or are badly needed in a production process. Which value one needs depends on how time gains are processed in the CBA. In the Netherlands, an investment in the transport networks is treated as leading in the first instance to time benefits only, but not to operating cost benefits; the time benefits are then monetised using a VTTS that includes the operating cost component and the cargo-related component.

5.7.3 Who should be interviewed – and what about?

In the 2004 freight VTTS survey, it was found that the road transport VTTS was between 80 and 98% of transport costs (as defined above in section 5.7.1), depending on the role of the good in the production chain (raw material or final product), the value of the good and containerisation. For rail this was 71%, for the water-based modes this was considerably lower. In the new Dutch VTTS and VTTR survey, there is an intention to disentangle the role of the transport services and of the goods (de Jong et al., 2007), on the basis of the following assumptions/hypotheses.

Table 5.5: Assumptions on freight VTTS (from de Jong et al., 2007)

	VTTS related to cargo	**VTTS related to vehicles and staff**
Carrier	0	~factor cost
Shipper that contracts out	Interest, deterioration, disruption of production, out of stock	0

Shippers with own account transport can give information on both the bottom-left and top-right element. Of course there may be exceptions to the above general pattern, but in the freight questionnaires, the Dutch study will steer the shippers that contract out to only answer the components they generally know most about (bottom-left), and likewise for carriers (top-right). If the new SP survey confirms the above hypotheses, then the mode-specific factor cost (or a time-dependent % of it) will be used, adding a commodity-based value for the VTTS related to the cargo.

5.8 Conclusions

Following from Chapter 4, which summarised evidence on VTTS for passenger, the intention of this chapter was to provide analogous evidence for freight. In comparison to passenger, the evidence base for freight is however extremely limited. A good reason for this is the difficulty of sourcing the requisite data, in sufficient volumes, and of adequate quality. This is an age old problem that continues to impede VTTS research on freight and logistics.

Even among the freight studies which have taken place, there would seem to be significant differences in method and context, making it difficult to draw clear conclusions based upon consensus. For this reason, inspection of „headline' estimates of VTTS does not often yield immediate insights, and it is not until one delves into the detail of how estimates have been derived that one can achieve the necessary level of understanding. At the detailed level, there exist a whole host of technical issues that could impinge upon our understanding of VTTS estimates, in particular its relationship with VTTR.

Against this background, we conclude by making two observations:

First, given the challenges of collecting fresh data (and the likely practical problem, in the context of any update study, of collecting data in sufficient volumes to reveal any new insights), we believe that there would be considerable merit in taking stock of the evidence that already does exist. This would however call for a detailed technical review of the evidence, so as to glean the necessary level of understanding.

Second, if fresh data were to be collected as part of an update study, then there is a case for carefully targeting those strands of the evidence base which are judged to be weakest. Moreover, we note that existing evidence is disparate, and suggest that resources might most productively be used to complete and synthesise the evidence base.

6 RESEARCH TASKS THAT WOULD BE INVOLVED IN COMMISSIONING OPTION A

6.1 Introduction

Following from the review activities described in Chapters 2-5, the next stage of the project involved consolidating observations from the review, and scoping up possible approaches to a potential study for updating recommended values of travel time savings in the UK. As outlined in Chapter 1, the Department distinguished between two possible approaches to such a study, namely:

- **Option A**: essentially to repeat the 1999 study, whilst taking advantage of recent developments in methods..
- **Option B**: this extends Option A, as necessary, to ensure coverage of „areas of concern' identified by the Department.

The present chapter deals with Option A, which would inform a possible update of valuations of non-working time per person, as currently documented in Table 2 of WebTAG Unit 3.5.6. In what follows, we specify two broad approaches to the implementation of Option A, thus:

- **Option A1**: In essence, this would involve a repeat of MVA, ITS & TSU (1987) and AHCG (1999), but qualifying that, with the benefit of hindsight, it may be unwise to repeat certain aspects of these studies and/or, with the passage of time, certain aspects may have become redundant.
- **Option A2**: This would be based on the same principles as A1, and encompass the same scope in terms of updating Table 2 of WebTAG Unit 3.5.6, but update the methods of MVA, ITS & TSU (1987) and AHCG (1999) to modern day standards.

We were advised by the Department that the study would not be required to deliver mode-specific VTTS, but should deliver VTTS that are transferable across modes. Noting this, we propose a „core' Option A1 study based entirely upon a survey of car drivers and passengers. This option is in line with the 1999 study, and would imply an assumption that VTTS for car drivers are, for purposes of appraisal, applicable to other modes (as was proposed, with some justification based on meta-analysis, in the ITS/Bates (2003) work). For both A1 and A2 we propose detailed study specifications, prescribing the phasing of the research, the scope of each phase (including modal coverage and recommended sample sizes), and the methods to be adopted in analysis.

We should caution that Option A embodies multiple dimensions. First and foremost, and with reference to the structure of Chapter 3, there is a distinction between survey methods and estimation methods; whilst these will need to cohere in order to deliver the requisite valuations, there are essentially separate sets of methods. Second, there is a question as to how much of the earlier surveys is worth repeating, particularly as, with the benefit of hindsight, we are aware of some aspects of the old surveys which must be corrected, e.g. the need to collect distance information. Third, within survey methods, there are a raft of more detailed issues to consider, relating to format, design, the survey vehicle, recruitment and locations of sites. These dimensions potentially give rise to a large number of possible study configurations, and make it difficult to set the options out wholly consistently. For these reasons, the specifications that follow should be taken as illustrative of a potential well-conceived approach to Option A rather than the single definitive approach. Moreover, it would be sensible to think of A1 as „loosely' repeating previous national VTTS studies, rather than slavishly following these studies to the letter.

In seeking to adhere, in a broadly faithful manner, to the methods used in AHCG (1999) and MVA, ITS & TSU (1987), a complicating factor is that these two previous studies show substantive differences in scope; whilst the 1987 study considered all modes, 1999 was restricted to road. More specifically, the 1987 study involved the following surveys:

- West Yorkshire Corridor: mode choice survey of car, bus and rail, using RP and Transfer Price (TP) methods.
- North Kent: mode choice survey of commuter coach and rail, using SP, RP and TP.
- M1 pilot: route choice survey of car, using SP.
- Exploratory study of private travel: in-depth survey of leisure travellers on all modes.
- Tyne crossing: route choice survey of car travellers, using RP and SP.
- Urban bus: service choice of commuting and leisure travellers, using SP.
- Long distance rail and coach: service choice of commuting and leisure travellers, using SP.
- Long distance car: route choice of commuting and leisure travellers, using SP.

By contrast, the 1999 study involved:

- Car driver and passenger survey: „value of time', „road characteristics', „departure time', „chance of delay', and „tolled vs. untolled route' games, using SP.
- Freight, coach and bus operator survey: within route and route choice games, using SP.
- Newcastle SP/RP survey: route choice survey of car travellers, using RP and SP.

Given the differences in the 1987 and 1999 studies, we shall here propose a synthesis, such that Option A1 involves a „core' car driver and passenger survey, and optional extras of an urban bus survey (Option A1.1) and long distance coach and rail survey (Option A1.2). The latter two surveys are drawn from 1987, and were not undertaken in 1999. These three surveys would together cover the key modes, for both short and long distances.

We have avoided the use of mode choice surveys (e.g. North Kent from the 1987 study), since these potentially introduce additional complications to the transfer of VTTS across modes. We have also omitted the freight, coach and bus operator survey (from 1999) and the Newcastle SP/RP survey (from both 1987 and 1999). These omissions are based on various considerations:

- First and foremost, the results of neither survey were taken forward when updating VTTS guidance subsequent to the 1999 study.
- Second, we propose alternative approaches to freight as part of Option B.
- Third, in the cases of coach and bus travel, we see no reason to depart from the current approach of valuing time savings using the „cost savings' approach for operators, and standard appraisal values (i.e. essentially a car value) for passengers.
- Finally, the recent M6 Toll study involved a comprehensive analysis of VTTS, and this would offer an appropriate analogy to a repeat of the Newcastle survey.

If Option A1 were to be commissioned without A1.1 and A1.2, then this would imply acceptance of the convention of appraising time savings for public transport at the same unit value as car; this was in essence the approach recommended by ITS/Bates (2003). Surveys of modes other than car would likely yield variation in estimates of VTTS by mode, leaving the (non-trivial) issue of reconciling these various estimates in order to arrive at a standard VTTS applicable for appraisal across modes.

Option A2 involves enhancing each component of A1, specifically:

- The focus group analysis is combined with a broader range of qualitative research methods
- The urban bus survey is updated to cover a broader range of public transport modes
- The long distance coach and rail survey is updated to include domestic air

We now consider A1 and A2 in more detail.

6.2 Option A1

Option A1 should be based primarily upon SP analysis, albeit preceded by focus group analysis. The reliance on SP is partly to ensure consistency with the 1987 and 1999 studies, and partly a reflection of our judgement that, at the present time, there is no alternative data source that could be considered a credible substitute for SP. In forming this view, we acknowledge that SP data has its weaknesses, and that alternative data sources - notably forms of RP data - may bring value. On balance, our view is that these alternative data sources should, where appropriate and practicable, be used not as a replacement for SP, but as a means of mitigating SP's weaknesses and thereby strengthening its validity.

SP experiments should be developed using established design methods/concepts used in previous VTTS studies, namely fractional factorial orthogonal design plans, boundary values, and the so-called „Bradley' approach (see Chapter 3). Indeed, where possible, researchers should employ direct repeats of the detailed experimental designs used in 1999 and 1987 (suitably updated for RPI). This is not to say that we necessarily recommend these methods/concepts, but their use would establish an objective basis for comparing 2010 valuations against 1999 and 1987.

We note that where old designs are resurrected, this may permit some economies in the design process.

6.2.1 Focus groups

In common with 1999, SP analysis should ideally be preceded by focus group analysis, covering two requirements:

- To develop understanding of the behavioural factors underlying the VTTS.
- To test whether the SP experiments, as designed, are understandable and meaningful.

Whilst it may be difficult to mimic in a precise fashion, it would be helpful if, as far as possible, focus groups could be conducted along similar lines to 1999. As well as informing the design of new SP experiments, this would hopefully provide a basis for comparing focus group analyses from 2010 and 1999. To this end, the analysis should involve 6 focus groups of 8 respondents, with 2 of these groups dedicated to each of the segments of business, commuting and leisure.

6.2.2 Car driver and passenger survey

Survey design

Car occupancy/group size can be a difficult issue to analyse. Whilst we would not go as far as to prescribe a specific „best practice' method, the chosen approach should certainly be clear and unambiguous as to whether VTTS relates to an individual or group. In the former, distinction should be drawn between driver and passenger. The latter should distinguish between different group compositions, e.g. adults vs. families with children.

To maintain consistency with 1999, this SP experiment should be implemented as part of a self-completion questionnaire, with postal return. Respondents would be recruited at filling stations. The selection of these filling stations should offer reasonable geographical spread across England, reasonable coverage of small and large cities, and avoid bias towards relatively expensive or inexpensive filling stations. Where possible, the same specific locations used in 1999 should be re-used, so as to eliminate any factors relating to location that could hinder comparability. Respondents should be subjected to a recruitment process which confirms that they are in scope, guards against non-response bias, and serves to identify the relevant SP questionnaire for that respondent. Participation could be incentivised through a prize draw or similar, provided that this practice does not unduly bias the sample.

Prior to implementation of a field survey, it is usually good practice to test the survey by means of a pilot. However, in the present case, it may be feasible (and indeed sensible) to recycle proven designs from 1987 and 1999, which may reduce the need for piloting. That said, it would seem prudent to conduct some degree of piloting, so as to reassure that historical designs continue to offer meaningful and relevant trade-offs. In order to minimise non-response bias, both in the pilot and field studies, respondents who fail to complete parts or the entirety of the survey should be followed up after a pre-specified period of time.

SP experiment

In order to maintain consistency with 1999, the basic format of the SP experiments should be a choice between a pair of travel options, bespoken to road type (i.e. motorway, trunk road, or urban) and distance. As noted earlier in Chapters 3 and 4, the 1999 study employed 5 distinct games, and only the first game - the „value of time' game - served to inform revisions to VTTS; we therefore focus Option A1 around this particular game. The „value of time' game describes travel options simply in terms of the attributes of travel time and travel cost. In the context of A1, it would be appropriate to issue respondents with several repetitions of this game.

We propose the sampling quotas detailed in Table 6.1. The quotas are roughly commensurate with those used in 1999 (albeit with some reallocation from business to non-work, on the basis that the focus of the present study is on non-work). Note that „urban' refers to A-roads, and „inter-urban' to motorways and non-urban trunk roads. Given the focus on non-working time, the „business' segment would not be required. There is however a question as to whether it would be most cost-effective to screen these out at the recruitment stage, or to include these as a segment of the analysis.

Table 6.1: Target sample quotas for car driver and passenger

	Business	Commuting	Essential (non-work)	Discretionary
Urban	500	500	500	500
Inter-urban	500	500	500	500
Total	1000	1000	1000	1000

Analysis of SP data

The raw data should be subjected to a cleaning/segmentation process, which would identify and treat any „irrational' or missing observations. Prior to estimation, there should be elementary analysis of the data. This should confirm the coverage of relevant sampling segments (e.g. personal disposable income, trip purpose, trip duration/distance), and analyse cross-tabulations against other travel behaviour segmentations, as well as socio-economic-demographic segmentations generally.

Following cleaning and elementary analysis, the data should be subjected to detailed estimation using logit methods. As a preliminary, separate models should be reported for key sample segments (e.g. income, trip purpose, trip duration/distance). Where relevant, such models should establish the prevalence of, and make appropriate provision for, „scale differences' across different segments/versions of the SP game, and „repeated observations' within experiments. Beyond the preliminary models, there should be investigation of the extent to which the models can be further segmented, guided for example by the segmentations investigated through cross-tabulation. More generally, there should be examination of the degree to which the VTTS distribution is identifiable from the data.

There should be consideration of the extent to which valuations of travel time savings show variability according to the following specific criteria:

- The absolute size of the change in travel time.
- The size of the change in travel time relative to the overall trip duration.
- The sign of the change in travel time.

There should also be analysis of variation in VTTS by:

- income
- car occupancy/group size
- the purpose of leisure travel (e.g. shopping, school journeys, social visiting, holiday, other leisure)
- socio-economic effects (e.g. occupation, age, gender, household composition)

More generally, there should be comparison between the principal VTTS estimates emerging from any update study and those reported by previous studies, namely MVA, ITS & TSU (1987) and AHCG (1999), as well as the standard valuations documented in WebTAG Unit 3.5.6.

6.2.3 Option A1.1: urban bus survey

Comment

This is a sector which has seen considerable investment, with the intention of achieving mode shift from car. The motivation for Option A1.1 would be to seek evidence supporting or refuting the proposition that car values are transferable to bus. More specifically, Option A1.1 would collate evidence on VTTS for bus users and investigate whether, once individual-level characteristics of travellers (which may well vary across modes) have been accounted for, VTTS shows variability according to different levels of comfort associated with different services.

Some SP evidence on VTTS for urban bus already exists (e.g. Hollander, 2006), and it may be sufficient to defer to this, if only to the extent of examining the relativity between values of time for car and bus. Such evidence could however be very case-specific (depending for example on the „quality' aspects of the particular public transport system being surveyed), making it necessary to base recommended VTTS for appraisal on some notion of a representative system or an average of valuations from a range of systems.

Survey design

Following the design of the 1987 study, this survey should be conducted in two stages. The first stage would involve an initial screening questionnaire, distributed and collected on board. The second stage would involve a self-completion SP experiment with postal return. The screening questionnaire should elicit information on origin and destination and journey purpose, as well as request a name and address for distribution of the SP. Ideally, the survey should be based on the same urban locations surveyed in 1987, namely Oxford and Leeds. As with the car survey, the bus survey should be piloted prior to field implementation.

SP experiment

In order to maintain consistency with 1987, the experiment should comprise 2 games, each involving the ranking of 10 „cards'; possible adjustments to this design are considered later under Option A2. The first game should be based on the following attributes:

- time to bus stop
- scheduled frequency
- fare

The second game should be based on:

- likely waiting time
- in-vehicle time
- fare

Sampling quotas

The MVA, ITS & TSU (1987) report is not explicit concerning target quotas, and we therefore propose quotas commensurate with the car driver survey (Table 6.2). Note that in this case the business segment is omitted at the outset.

Table 6.2: Target sample quotas for urban bus survey

	Commuting	**Essential (non-work)**	**Discretionary**
First game	500	500	500
Second game	500	500	500

6.2.4 Option A1.2: long distance coach and rail survey

Comment

Analogously with the bus survey (Option A1.1), this option would investigate discrepancies between VTTS for car and public transport, but this time for long distance as opposed to short distance trips.

Survey design

This survey seeks to replicate the long distance public transport mode choice survey of the 1987 study, and to that end should be conducted on a route or routes offering choices between coach and rail. Ideally, these routes would be the same as those surveyed in 1987, namely London-Leeds, London-Bristol and Manchester-Birmingham. Whilst the survey would not explicitly offer a mode choice, the existence of different journey alternatives would give realism to an SP experiment.

The survey should be conducted in two stages. The first stage should involve an initial screening questionnaire, distributed and collected on board. The second stage should involve a self-completion SP experiment with postal return. The screening questionnaire should elicit information on origin and destination and journey purpose, as well as request a name and address for distribution of the SP. As with urban bus, business travellers should be screened out. As before, the survey should be piloted prior to field implementation.

SP experiment

Maintaining faithfulness to the 1987 study, the SP experiment should involve a comparison between 12 pairs of journeys, with respondents asked to indicate a preference according to a 5-point scale. We will review this basic format under Option A2.

Each journey should be described in terms of:

- fare
- scheduled journey time
- maximum delay
- a timetable showing scheduled departure and arrival times for a range of services

Survey quotas

In the absence of a clear steer from the MVA, ITS & TSU (1987) report, we propose the sampling quotas given in Table 6.3. We specify distinct quotas for both coach and rail users; segmentation by journey purpose would be less appropriate for this survey, since business is out of scope, and the long distance commuting market will be relatively small.

Table 6.3: Target sample quotas for long distance coach and rail survey

	Coach	Rail	Total
Inter-city to London	500	500	1000
Inter-city non-London	500	500	1000
Total	1000	1000	2000

6.3 Option A2

Whereas Option A1 sought to replicate the MVA, ITS & TSU (1987) and AHCG (1999) studies in a generally faithful manner, the purpose of Option A2 would be to update methods to present day standards and conventions. To this end, we recommend that A2 should proceed on two fronts. First, we would recommend various enhancements to methods generally, and that these should apply to all options commissioned. Second, we would recommend specific enhancements to some surveys. In what follows, we will outline the generic enhancements in the context of the „core’ car driver and passenger survey, before considering specific enhancements to the urban public transport survey (Option A2.1) and long distance public transport survey (Option A2.2).

6.3.1 Car driver & passenger survey

Survey design

Sampling methods should seek to achieve representativeness against pre-specified stratification (e.g. by region), and relevant segmentations (e.g. such as those adopted by NTS). Specific quotas should be set with reference to those sub-samples. In analysis, data should be weighted as necessary to account for any bias that might arise during the process of collection, e.g. non-response. As was shown by ITS/Bates (2003), failure to account for such influences could lead to biases in estimates of VTTS.

Whereas Option A1 was reliant upon pen-and-paper questionnaires, Option A2 should consider possibilities for alternative methods of survey administration, such as computer-assisted and web-based questionnaires. The crucial advantage of the latter methods is that they allow for customisation of the experiment. Whichever survey method is adopted, the analyst should explain the rationale behind that decision, and take appropriate measures to guard against/mitigate any associated bias in the data sample, for example against participants without access to/unfamiliar with the internet. A particular need is to ensure that questionnaires are meaningful and relevant to respondents, and this should be promoted through appropriate design and administration of surveys.

SP experiment

SP design methods have advanced since the 1999 study, although we observe that this remains an evolving research area. As was discussed in section 3.2.4 above, a promising development is the use of so-called „efficient' SP design methods, which have been applied to recent VTTS studies in some countries (notably Australia). The power of efficient design methods is in maximising the informational content of data from a given respondent, potentially reducing the overall number of respondents (and cost) involved in data collection. We envisage a possible role for these and other emerging methods within Option A2. Whilst such methods may, in principle, bring benefits, analysts should be careful to benchmark against more established methods, for two reasons. First, it would help to build confidence in the validity of the emerging methods. Second, it would provide a basis for comparing 2010 results against 1987 and 1999.

Again with reference to section 3.2.4, the recommendations of AHCG were based upon an „abstract' binary choice experiment trading time against cost. Option A2 should consider the scope for a broader range of experimental formats (for example, with more than two alternatives, and/or with „labelled' alternatives), since there is evidence that valuations may be sensitive to format.

Sampling quotas

The preceding points may have implications for the sample quotas detailed under Option A1, and we therefore leave it to the judgement of the analyst to propose target quotas. As well as taking accounting for the survey/SP methods adopted, quotas should be further determined with reference to NTS. Moreover, the analyst should present a clear justification - in terms of statistical theory - for the target quotas proposed, ensuring that, irrespective of the methods adopted, the informational content of the data will, at worst, be commensurate with the quotas detailed under A1.

Analysis of SP data

With reference to earlier discussion on cleaning in section 6.2.2, the analyst should review and reconsider procedures adopted in „cleaning' the data, since the procedures adopted could have a significant impact on the results obtained. Further advice on this is given in WebTAG Unit 3.11.3.

Having developed standard logit models, analysts should consider the scope for more flexible models, especially those from the mixed logit family (e.g. as documented in WebTAG Unit 3.11.5), but also potentially latent class methods, and non-parametric methods. Note that non-parametric methods could potentially guide the specification of parametric methods. In particular, analysts should seek investigate the prevalence of taste variation within pre-defined segments (e.g. relating to income, journey purpose, etc.), and any bias and inefficiency associated with repeated observations.

Further analysis might follow the precedent of the Danish VTTS study in modelling SP data on WTP space (i.e. in terms of βtime/βcost) rather than utility-space (i.e. βtime or βcost). An attraction of this approach is that it allows direct revelation of the VTTS distribution, and a means through which the analyst can verify that the distribution is identified. The WTP approach also brings convenience when analysing reference dependence, as might apply to STS for example.

If (as we recommend in due course) both Options A1 and A2 are commissioned, then the analyst should first estimate separate models on data from each option. The analyst should then explore the potential for estimating models on the merged data, since this would serve to maximise the informational value of the dataset. Where there is likely to be an interest in merged models, it is sensible to anticipate this need beforehand, since this could influence the manner by which the SP experiment is designed. Furthermore, opportunities should be explored for reanalysing historical data collected for MVA, ITS & TSU (1987) and AHCG (1999) using current modelling methods, and for developing models on fresh data merged with the historical data.

6.3.2 Qualitative analysis

Analysts should consider opportunities for combining the focus group analysis of Option A1 with a broader range of qualitative analyses. Of particular importance is to make sure that the SP experiment is relevant, understandable and realistic to the respondent. To these ends, the analyst could usefully employ techniques such as cognitive interviewing. This would seek to achieve a deeper understanding - at the level of the individual traveller - of how the SP experiment relates to the respondent's personal travel experience and, more generally, of how valuations of travel time and travel time reliability might impinge upon their travel behaviour.

6.4 Option A2.1: urban public transport survey

In updating the urban bus survey, we observe that public transport opportunities in some locations have changed markedly over the last 20 years. We therefore propose that Option A2.1 should again be a within-mode survey, but administered across a broader range of PT modes potentially including conventional bus, high quality bus and LRT.

Analysts may wish to reconsider the earlier recommendation (section 6.2.3) for a ranking experiment. There may be attractions, from the viewpoints of coherence and economy, in employing a common survey vehicle across experiments. Such attractions might however be traded-off against the possibility that different survey vehicles could elicit difference behavioural insights. Indeed, it occurs to us that, compared to the car driver and passenger survey, a ranking experiment would make the public transport survey relatively complex.

Moreover, the approach adopted by MVA, ITS & TSU (1987) was not without its complications, in particular due to a) trying to modify access time without very good justification (e.g. moving bus stops) and b) specifying wait time, without getting involved in timetables and adherence to schedules. For these reasons alone, the basic approach of the survey might benefit from review and revision.

6.5 Option A2.2: long distance public transport survey

The key developments in this sector are the advent of domestic air, and the policy interest in high speed rail. Updating to 2010, we propose that Option A2 should be conducted on a route or routes offering choices between rail and coach or air.

In a similar vein to the discussion at the outset of section 6.2.3, we acknowledge that, instead of commissioning fresh SP analysis, existing evidence (e.g. in relation to HSL) could potentially be exploited. The attraction of fresh SP analysis would be to ensure coherence with the methods used for car, as well as generally adding to and reinforcing the evidence base (e.g. as was the case with Atkins' SP work for Planet Strategic). Analysts may wish to reconsider the earlier recommendation for a preference scale experiment, again trading off the possible attractions of a consistent survey vehicle with the possibility that different vehicles may reveal different insights.

6.6 Conclusions

This chapter has scoped out research studies within the framework of Options A1 and A2, wherein:

Option A1 would involve a (partial) repeat of MVA, ITS & TSU (1987) and AHCG (1999).

Option A2 would update the methods of MVA, ITS & TSU (1987) and AHCG (1999) to current day standards.

Our proposals for Option A are summarised in Table 6.4 below.

Table 6.4: Basic structure of Option A

Option	CORE		ADD-ONS	
A1	Focus groups (1999)	Car driver & passenger (1999)		
A1.1			+ Urban bus (1987)	
A1.2				+ Long distance coach and rail (1987)
A2	Qualitative research (2010)	Car driver & passenger (2010)		
A2.1			+ Urban PT (2010)	
A2.2				+ Long distance PT (2010)

Notes:
„Car driver & passenger (1999)' refers to a repeat of the car driver and passenger survey from AHCG (1999); „Car driver & passenger (2010)' refers to a car driver and passenger survey using current best practice; „Urban bus (1987)' refers to a repeat of the urban bus survey from MVA, ITS & TSU (1987), and so on...
+ denotes an appendage to the „core' study, where the core comprises focus group or qualitative research and a car driver and passenger survey

We believe that there are attractions in both strands. On the one hand, Option A1 would (with appropriate controls on econometric methods) provide a means of objectively analysing whether VTTS has changed over time. On the other hand, some aspects of 1987 and 1999 methods have, with the benefit of experience, been found to have limitations; in such instances there is good reason to employ newer and/or more defensible methods. It seems to us that, should the Department commission either of A1 and A2 on their own, then they would leave themselves open to the accusation either that methods underpinning VTTS estimates were deficient (if A1 were commissioned), or that VTTS estimates from 2010 were not comparable with 1999 (if A2 were instead commissioned).

We therefore believe that there is a very strong case for combining elements of A1 and A2. By marrying historical methods with current best practice methods, A1+A2 would maintain consistency with 1987/1999, and (in our view) minimise the overall level of methodological risk. This would however have cost implications relative to commissioning A1 or A2 individually, making it necessary to form a judgment on the acceptability of the additional costs against the analytical benefits. We will consider such issues in Chapter 11.

7 RESEARCH TASKS THAT WOULD BE INVOLVED IN COMMISSIONING OPTION B1 (ON 'THE VALUE OF RELIABILITY')

7.1 Introduction

Reliability (better, Travel Time Variability or TTV) is an increasingly important attribute, and its importance is likely to increase further as long as demand is growing and investment is in short supply. In line with this, DfT has recently issued draft guidance (WebTAG 3.5.7) on how it should be accounted for in appraisal. However, it is worthwhile pointing out that there remain practical difficulties relating to the **forecasting** of changes in reliability (this emerged clearly from the recent TRB conference in Vancouver). As far as the current project is concerned, interest in TTV relates to its valuation (VTTR), but clearly the useful implementation of any improved values will require progress on the supply front to be made. In other words, VTTR is only a useful concept if model outputs are available to apply it to.

In addition to the interest in VTTR *per se*, there are also concerns, though perhaps of lesser importance, that without a careful specification the standard value of time (VTTS) could be "contaminated". For example, the 1980s UK study suggested that VTTS was higher in congested conditions, and that this might reflect increasing unreliability.

The guidance in 3.5.7 recommends the use for highway appraisal of the standard deviation as an index of TTV, and the valuation is based on the concept of the "reliability ratio" ρ. This is the ratio of the disutility of one unit of standard deviation to the disutility of the corresponding unit of travel time. Then, if v is the VTTS, the value attached to an improvement in the standard deviation of one unit is equivalent to $\rho.v$. The current generally recommended value for all (personal) travel purposes for ρ is 0.8, a value essentially derived from the Cranfield work in 1993, though it has been supported by some other work since then.

Apart from a few specialized RP studies in the US involving tolled lanes, valuations of TTV have been entirely obtained from SP. Nonetheless, there remain serious problems relating to the presentation of TTV in SP experiments. Recent work for the Dutch Ministry of Transport by Significance has established what is probably the state of the art in this respect. The proposed surveys, which are about to begin, contain three separate SP experiments:

- SP1: tradeoff between time and cost only, to allow the estimation of VTTS, using the formulation $U = \lambda.C + \alpha.t$, whence VTTS = α/λ

SP2 and SP3 have the same presentation format, which gives a specified departure time, a set of possible travel times, and cost. The difference between them is that in SP3, the departure time is set so that the expected arrival time is in fact fixed. Thus SP3 has a restricted set of possible models.

- SP2: allows the estimation of the value of both mean travel time and the standard deviation σ, as well as a “scheduling” formulation (valuation of early and late time), using the formulation $U = \lambda.C + \alpha.E[t] + \theta.\sigma + \beta.E[SDE] + \gamma.E[SDL]$, where SDE and SDL are the early and late “schedule delay” (relative to the preferred arrival time).
- SP3: allows the estimation of the value of both mean travel time and the standard deviation, using the formulation $U = \lambda.C + \alpha.E[t] + \theta.\sigma$, whence VTTS = α/λ and $\rho = \theta/\alpha$

Substantially the same experiment is currently also being used by the Norwegian Transport Economic Institute (TØI).

SP1 is the “standard” approach for VTTS, similar to the AHCG SP1, while SP2 and SP3 introduce the concept of TTV. SP3 corresponds closely with the approach taken in the 1993 Cranfield work, and, in comparison with SP1, allows the testing of whether the VTTS is affected by the explicit presentation of TTV. SP2 allows the equivalence between the “mean/sd” formulation and the “scheduling” formulation for TTV to be tested, at the cost of greater complexity for the respondent, as well as testing whether making scheduling implications implicit influences the reliability ratio.

As we discuss further below, it has often been observed that, having re-optimised departure time to deal with TTV, the sum of the additional scheduling terms ($\beta.E[SDE] + \gamma.E[SDL]$) is more or less linear in sigma. This empirical observation has recently been proved more generally by Fosgerau & Karlström (2009) [F&K]. If this holds good (at least approximately) then more or less the same level of explanation will be provided by including a) $\theta.\sigma$ and b) the sum of the ESDE and ESDL terms. Hence, the theoretical expectation in respect of SP2 is that there will be some redundancy in the utility formulation.

As noted, the current guidance implies that the reliability ratio is constant for all purposes, which hardly seems satisfactory: in addition, since the valuation of TTV is essentially due to timing constraints, the factors leading to variation in the reliability ratio could be rather different from those that lead to variation in VTTS. Any empirical investigation should allow these potentially different dimensions to be determined.

Of course, there will still be practical issues as to how far standard transport modelling will support the distinctions which might emerge (for example, concerts and theatre with fixed start times, as opposed to going to the gym).

7.2 An alternative approach

In spite of the considerable effort put into the presentation of TTV in SP experiments (see, e.g. Tseng et al, 2008), there is concern that the respondent burden may be too great. On the one hand, it is necessary to ensure that the respondent grasps the basic notion of uncertainty, and the possibility of different outcomes, but on the other hand, it is likely to be too burdensome to present the distribution in adequate detail. It must also be remembered, of course, that the respondent is actually being asked to choose between (at least) two alternatives having different levels of TTV.

The current "standard" (and that adopted in the Dutch study) is to present five equi-probable journey times. However, while this may convey the notion of TTV to the respondent, it is too coarse to give any idea of the shape of the distribution, or even an accurate measure of σ.

Given these concerns, it has been recently proposed by Bates (2009) that it might be preferable to concentrate on SP experiments which present scheduling information **without** uncertain travel times, and then use theoretical models (in line with Fosgerau and Karlström, 2009), together with empirical data on TTV, to convert to the reliability ratio.

Assuming a "standard" scheduling utility function of the kind proposed by Small/Vickrey to model departure time choice under conditions of travel time certainty:

$$U = \alpha.t + \beta.SDE + \gamma.SDL$$

Fosgerau shows that the existence of TTV represented by a distribution f(t) for travel time leads to an expanded disutility U(f), say, which can be written as:

$$U(f) = \alpha \cdot E(t) + (\beta+\gamma) \cdot H(\text{standardized } f, \beta, \gamma) \cdot \sigma$$

whence the VTTR (value of travel time variability) is $(\beta+\gamma) \cdot H(\text{standardized } f, \beta, \gamma)$

The function H is known from theory, or can be derived by numerical analysis.

The values of α, β and γ could be derived from an SP experiment which merely traded off different departure/arrival times with different (certain) travel times (and perhaps money costs). Applying the Fosgerau approximation, this could therefore produce usable values of VTTR without the need for explicit presentation of TTV.

It should be noted, however, that this approach is not universally accepted: in particular, proponents of experimental economics lobby would probably not subscribe to the proposition that preferences are stable across riskless and risky domains.

7.3 Further issues

There are additional complications in the case of public transport. This discussion focuses mainly on the car mode, and but we note here some of the key differences raised by public transport.

The first point relates to VTTS, as opposed to TTV. Generally, the effect of TTV will be to increase the mean travel time, compared with the free-flow travel time on the highway, or the scheduled time on pt. In the case of highway time, this increase is partly catered for, by means of the speed-flow relationship: it may be assumed that this takes account of the average impact of day-to-day variability, though **not** of incidents (which is the reason why the INCA program calculates an additional benefit due to mean time as well as TTV). In the case of rail, on the other hand, time benefits are, as standard, calculated on the scheduled times, so that – quite apart from the effects of TTV *per se* – there is a need to reflect the additional mean delay.

The current PDFH recommendations on rail variability, which have been incorporated into WebTAG 3.5.7, are in line with work reported in Bates *et al* (2001). This found that mean delay time, over and above the advertised schedule, was perceived as much more onerous than scheduled time, suggesting an average of 2.5 times the standard VTTS. Given this, the **additional** effect of TVV *per se* appeared quite small, especially because the existence of fixed timetables implied that, in terms of departure time choice, travellers were already incurring substantial schedule delay. On this basis, the work suggested that the effect of TTV might be taken account of by increasing the delay multiplier from 2.5 to 3.

Recent work for the DfT by ITS on rail reliability has provided evidence of a slightly different kind. Further work is required to see to what extent the ITS work supports or challenges the PDFH approach.

In the case of the bus mode, the impact of **waiting** time is likely, if anything, to be more serious than in the case of rail. There is little evidence of recent work on bus reliability, apart from Hollander (2007) and a FIT project at ITS Leeds with the title "Model To Assess Public Transport Reliability" [see Liu, R. and Sinha S. (2007)].

7.4 Implications for current study

7.4.1 Options B1.1 and B1.2

In terms of reliability, one option is to proceed along the lines of the Dutch and Norwegian studies, which could be adapted without significant difficulty, especially for the passenger surveys (the freight surveys could be more complex and these are discussed elsewhere). We set out more detail about this option below.

Nonetheless, it might be more prudent to wait for the outcome, which should be available within a year. If either or both of the Dutch and Norwegian studies may be considered a success, in terms of delivering implementable results and demonstrating internal consistency in the SP data, then either the essential results (e.g. in terms of the reliability ratio, in the case of the car mode) could be **transferred** (which we refer to as Option B1.1), or the study could be repeated for the UK (Option B1.2). Although there would probably be some preference for having UK-specific values, a pure transfer could be justified, particularly if the Dutch analysis appears well-founded. In any case, this could be revisited once the Dutch results were available: preliminary results for passenger VTTR (and VTTS) are likely to be available by summer 2010, with final results following in the autumn.

However, since there are **some** advantages in dealing with VTTS and VTTR simultaneously, the implications for the overall VTTS study need to be carefully considered. There are a number of issues involved. On a technical front, the question is whether we think that VTTS experiments are 'pure' – i.e. no cross-contamination from reliability. We suspect that the answer is: no more and no less than in 1987 or 1999. Nevertheless, given the increased importance of TTV, it might now be considered necessary to resolve this.

More generally, it depends on what is seen as important for VTTS. If the aims do not go much beyond Option A, then this would tend to imply a less ambitious approach to VTTR. By contrast, if the aims for VTTS are greater, then there will of necessity be a longer preparation period before the main surveys commence, and adopting a waiting policy on the Dutch/Norwegian work would probably not greatly affect the timescales. This would also afford the opportunity for a small amount of work to be put into testing the alternative approach discussed above – the scheduling approach under **certain** travel times within an SP context. This would provide a potential fallback in the event that the Dutch/Norwegian studies are not successful.

Clearly there are beneficial survey implications from bundling VTTS and VTTR together in a single programme contract. There may also be general aspects of consistency and continuity of analytical effort that are important. Yet another issue is whether – if a „wait and see' approach was taken to VTTR – it would be necessary to make earlier (and hence independent) progress on VTTS, or whether some delay would be acceptable?

The broad tactical possibilities seem to be as follows:

Option B1.1: Decouple VTTS and VTTR: proceed with VTTS and wait for Dutch results on VTTR, which – if successful – could be simply transferred. If transferring was not acceptable, or if the Dutch VTTR was not successful, it could then be decided whether additional survey work was appropriate

Option B1.2: Delay whole study until Dutch results are available (though some preliminary work could proceed). Then, if Dutch VTTR study is successful, a joint VTTS/VTTR survey programme could be launched. If Dutch VTTR study is not successful, it may still be possible to launch a joint programme, but using the alternative approach for VTTR.

Given that, at least on the passenger side, the delays will not be substantial, our recommendation is for Option B1.2.

In considering the best approach, it is clearly useful to sketch out what kind of effort is likely to be required to (more or less) repeat the Dutch work in the UK context. There follows a résumé of the Dutch approach. We concentrate here on person travel, though it should be noted that Freight travel is also included.

The design both for the sample, questionnaires and the SP experiments was carried out in the previous 2007 study. The main study consists of 5 phases:

- Preparation
- Pilot study
- Main study
- Data analysis

- Final report

The aim of the Preparation stage is really to get practical “buy-in” to the questionnaires from a number of organisations that are involved in passenger [and freight] transport, possibly leading to textual changes, which would be incorporated in the final versions of internet- based and CAPI versions of the questionnaires.

Phase 2: Pilot

The aim of the pilot is to test not only the questionnaires but also the fieldwork procedures for the main phase. To this end, the data collection will follow exactly the recomendations for the main survey. The Dutch pilot study will include 100 car drivers, 50 train-/metro travellers, 50 bus-/tram travellers, 50 travellers on high speed rail and plane and 25 recreational boat users. [In addition 25 organisations will be surveyed for freight transport.] Note that all person travel will be drawn from an internet panel (which contains 240,000 participants, from whom a 40% response rate is expected).

Phase 3: main survey

In total at least 5200 successful interviews will be carried out for person travel [and 520 for freight]. The planned segmentation is given in the tables below (referred to as Tables 14 and 15), which are reproduced directly from the Dutch scoping study report (de Jong et al., 2007). Within each segment for **person** travel the sample will be nationally representative in terms of income, employed status, age, sex and region.

Given that the interviews for person travel are self-completion, help facilities will be provided both in terms of a telephone helpdesk and computer-based aids.

Phase 4: Data analysis

This takes place in four stages:

Stage 1: Estimation of a base model, using only the attributes presented, separately for each purpose and mode, followed by limited investigation of further segmentation. These models will be based on multinomial logit. This will also allow comparison with previous models estimated on similar datasets.

Stage 2: This will involve alternative model formulations, including scheduling terms, non-linear effects (including gains and losses), and inertia.

Stage 3: This will bring SP and RP data together in a combined analysis.

Stage 4: This will include some investigations using mixed logit.

The sample sizes proposed are set out in Tables 7.1 and 7.2, for person and freight travel.

Table 7.1: Proposed sample sizes for Dutch passenger survey

Experimental segments	Target segments		Targets
Passenger private car transport	Trip purpose and time of day	Commuting Business Others Peak, off-peak	2500 for car drivers in total. with 1250 for commuting 500 for business 750 for others 50% peak / 50% off peak
Passenger public transport	Mode, trip purpose and time of day	Train/metro, bus/tram, high speed rail, air Commuting Business Others Peak, off-peak	2500 for passenger public transport in total. with 1250 for commuting 500 for business 750 for others Among these trip purposes, 40% is for train/metro/lightrail/“sneltram”, 40% is for bus/tram, 20% is for high-speed rail/air. 50% peak / 50% off peak 50% with transfers within public transport (e.g. bus-train, train-train)/50% without
Recreational navigation	-	-	200

Table 7.2: Proposed sample sizes for Dutch freight survey

		Road	Inland waterways	Sea	Rail	Air
Container	**Carrier**	50	40	20	20	
	Contract out shipper	50	20	30		
Non-container	**Carrier/own account shipper**	50	40	30	30	30
	Contract out shipper	50	20	20		20
Total		200	120	100	50	50

Clearly, these samples could be reviewed. It is unlikely that recreational navigation is a significant topic in the UK. In addition, it remains an open question as to whether the study should include public transport, while the current proposal is to exclude business travel. However, while this would reduce the number of categories to be explicitly investigated, it would be unwise to reduce the implied sample in each category. Indeed, if there is some interest in regional variation, this would imply somewhat higher sample numbers per category.

It may be noted that the first two phases are relatively light, involving perhaps 20 man-days each (NB exclusive of fieldwork costs), over a period of 5 months. Hence, on the assumption that the same survey procedure would be followed, the set-up costs are not high.

7.5 Requirements for appraisal

We have noted the practical difficulties relating to the **forecasting** of changes in reliability. In the interim it is probably adequate to continue to use the value of ρ = 0.8, as suggested in current Guidance (TAG 3.5.7).

On the supply side, the Guidance notes [para 2.1.11]:

"What is therefore required on the highway side to assess reliability is:

- *to model the impact of incidents on average journey time, and*
- *to model the level of variability (remaining with the approach of representing this by the standard deviation of travel time) associated with both incidents and DTDV.*

"How this can be done in practice is discussed in Section 3."

In following sections, the advice is given that for motorways and dual carriageways, Incident Cost-benefit Assessment (INCA) should be used; for Urban roads, the formulation developed by Arup (based on London and Leeds data) should be used; for single carriageways outside urban areas there is no empirical relationship for σ (the "Stress" method is recommended); for public transport, no method is given (though the approach to valuation is discussed).

A number of comments may be made on this position. Firstly, the INCA program operates on a very small scale, and while the methodology could, in principle, be generalised to a complete network, no such attempt has been made in practice. It may be worth quoting from the INCA Manual in this respect:

"Typical schemes for which the INCA methodology is appropriate range from motorway traffic management measures to motorway widening schemes which change incident-related delays and travel time variability."

INCA is thus essentially a tool dealing with a limited sequence of grade-separated links, with a correspondingly limited set of demand flows. Only 63 links (including "feeder links") and only 63 [or 16??][9] movements (which need to be pre-assigned to links) are catered for. There remains the question, in any case, as to what should be done about other links in the network.

With regard to the "Urban" highway case, it is in principle possible to apply the recommended relationships at the matrix level – essentially as a post-processing option: there is no experience of the convergence properties of the supply-demand loop when this is done. Note also that updated relationships have been derived (though of essentially the same form) in recent work by Hyder: these have not yet been incorporated into the Guidance.

[9] The INCA documentation is ambiguous – both numbers are mentioned.

Finally, in the context of multi-modal models, there is no guidance as to what should be done about public transport.

Hence, even if trusted and approved results were available for the **valuation** of TTV, which could also be included within demand models as a component of generalised cost, there is still a reasonable amount of work on the supply side to be done before this could be implemented in an overall modelling system.

The following may also be noted. Empirical investigations of factors affecting the standard deviation of travel time show a strong relationship with the (mean) travel time itself: in other words, highway congestion affects both mean travel times and the standard deviation, essentially because of high volume/capacity ratios. Many schemes which address either capacity or demand will tend to impact on both the mean and the sd of travel time. In such cases, a simplified approach would be reasonable.

Suppose as an approximation that $\sigma = a + \lambda(t - t_{FF})$, where t is travel time and t_{FF} the free-flow time. If this linear relationship holds before and after the scheme, we expect $\Delta\sigma \approx \lambda.\Delta t$. The overall valuation of the savings in both travel time and TTV is then given by $v.\ \Delta t + \rho.v.\ \Delta\sigma \approx v.(1 + \lambda.\rho).\Delta t$. Thus reliability could, in such simple cases, be included in the CBA by "loading" the standard VTTS by the factor $(1 + \lambda.\rho)$.

Clearly this would not be appropriate for schemes which were expected to have differential impact on TTV (for example, imposing speed limits to ensure a more even pattern of flow). However, such schemes might well be more the exception than the rule.

7.6 Possible short term actions

7.6.1 Option B1.3

Although it has been argued that there is wisdom in postponing empirical SP work on TTV until the results of the Dutch/Norwegian studies are known, if it is nonetheless decided to proceed, we believe that there is relatively little work to be done to convert the existing Dutch approach to something suitable for the UK environment. We would recommend the commissioning of a short „think piece' to work through the procedures that would be involved in translating methods (Option B1.2), but especially evidence (Option B1.1), from The Netherlands to the UK.

7.6.2 Option B1.4

On the alternative assumption that VTTR work would be postponed, it is recommended that trials are carried out of the "alternative approach" using a scheduling SP without uncertain travel times. This would be mainly new work, though one could look to some existing SP data on departure time choice (e.g. that investigated for time of day choice). The previous work (e.g. Hess, Daly & Bates (2005), "Departure Time and Mode Choice: Analysis of Three Stated Preference

Data Sets", Report to DfT) was aimed more at macro-time choice (though it was originally analysed as for a micro- context).

Such an option should be viewed as added security, and is being recommended on the basis that the Dutch VTTR work is by no means guaranteed to be successful. It needs to be accepted that, if the Dutch work **was** successful, it could be argued that this preparatory work had been wasted.

Outside the scope of this study, we would recommend that DfT take urgent action to address some of the existing "gaps" in the supply side of TTV, at least on an interim basis, though we would expect that this is already included in DfT's research agenda. Appropriate empirical evidence on the scale of TTV should also be collected, both to support the research agenda and because the F&K method (as noted above) requires a distribution of TTV, which should preferably be empirically based. There is existing data for rail and air (though it could take some effort to get it into the right format), and lots of electronic data for highway, though a lot more clarity is needed on how to use it.

7.7 Conclusions

Drawing the above discussion together, we feel that there would be considerable merit in seeking to make best use of insights from the current Dutch study of VTTS and VTTR. Even if this broad strategy were to be followed, there does however remain a tactical question of whether to simply transfer estimates of VTTR from The Netherlands (Option B1.1), or instead commission new estimates of VTTR for the UK but using methods developed in The Netherlands (Option B1.2). On balance we would opt for the latter, since this would take account of possible confounding of VTTS and VTTR, and also avoid any potential criticism regarding the „sovereignty' of VTTR estimates. A possible downside of Option B1.2 is that this would call for a delay in the commissioning of the whole VTTS update study, since VTTS and VTTR would need to be commissioned together, and the specification of the study would be contingent on the outcomes from the Dutch study. By contrast, Option B1.1 could proceed immediately, and combine estimates of VTTS from the UK with VTTR estimates from The Netherlands as and when the latter become available.

In order to inform tactics, there would be some merit in first commissioning a think piece (Option B1.3) to work through the procedures involved in transferring methods and (especially) evidence from The Netherlands to the UK. On completion of the think piece, the Department would be better informed to choose between Options B1.1 and B1.2.

Irrespective of which option is settled upon, it would be prudent of the Department to additionally commission a pilot study of the „alternative' approach to estimating VTTR (Option B1.4). The success of the Dutch study is by no means guaranteed, and this pilot would serve to establish a contingency.

8 RESEARCH TASKS THAT WOULD BE INVOLVED IN COMMISSIONING OPTION B2 (ON 'SMALL TIME SAVINGS' (STS))

8.1 Introduction

The purpose of this chapter is to outline specific options for the conduct of Option B2, which concerns the analysis of small time savings in the context of an update to appraisal values for VTTS. In what follows we propose three such options, specifically:

- Option B2.1: Extensions to analyses conducted as part of A1 and A2
- Option B2.2: Review of theory, evidence and practice
- Option B2.3: Exploratory development of new methods

Before proceeding, we should clarify the focus of our discussion. As will already be apparent from the earlier review chapters, the size of time savings is often considered in conjunction with sign of time savings. In what follows we focus specifically upon size as distinct from sign, justifying this on the basis of the findings of ITS/Bates (2003). The latter concluded that: *„we believe that the AHCG conclusion relating to significant differences in valuation according to the sign of both time and cost changes is invalid, due to a model specification error. This in turn relates to that part of the SP design which allowed direct comparisons with the "current journey". Although in our view it would be better not to include such comparisons, it is possible to make an appropriate allowance for them in the model specification. When this is done, the "sign effect" effectively vanishes'* (p25).

8.2 Option B2.1

8.2.1 Background

Option B2.1 would involve a moderate bolstering of the methods outlined under Options A1 and A2. That is to say, B2.1 would broadly follow the analyses of STS conducted as part of the 1987 (which was more theoretical than empirical) and 1999 (which did include empirical work) studies, but strengthened in places to account for the findings of ITS/Bates (2003) and provide for more extensive qualitative research.

The purpose would be to strengthen the SP in relation to STS and revisit the problem area encountered by the 1999 study. One idea worth pursuing would be to experiment with alternative framings, for example WTP for a STS on a one-off basis versus WTP for a STS on a recurrent basis, for example on a commuter rail trip when the train is speeded up. Closer engagement with TfL, in the context of their experience with London Underground appraisals, would also be in order.

The outcome of this might be some reinforcement of the current policy position, but would be unlikely to provide enough evidence on its own to support a move away from the status quo. Other similar ideas might include a revisiting of the logical underpinning of the status quo (the „egg box' analogy appears to have found favour) and a review of the statistical properties of the time savings themselves-should there be a test of statistical difference from zero and is simple summation of time saved an unbiased estimator of quantity (as opposed to value)? Another possibility would be a „think piece' on the form in which the Department might best deliver on its commitment to publish the size distribution of time savings.

Our overall view is that the „Common Unit Value' (CUV) assumption is rather like old age - „not so bad when you consider the alternatives'. We are struck by the relative lack of conviction in defending alternative positions, although it might be useful to articulate what the „best' alternative would be and consider its properties.

8.2.2 Tasks involved in B2.1: Extensions to analyses conducted as part of A1 and A2

The focus group analysis of Option A1 would be broadened to include a dedicated strand aimed at investigating the prevalence of „thresholds' and/or „budgets' inherent within travellers' decision-making, especially in relation to travel time. Such thresholds/budgets might take on various dimensions, such as the timetabled journey time, the free-flow journey time, the expected journey time (accounting, for example, 7for day-to-day variability), the „preferred arrival time', and so on. The intention of this analysis would be to reveal insights that would inform the specification of models described in the following points.

Modelling would broadly adhere to „conventional' analyses of STS, using the methods employed in AHCG (1999) and ITS/Bates (2003), but applied to fresh data from Options A1 and/or A2. Mindful of the so-called „inertia effect' identified by ITS/Bates (2003), analysts should, where applicable, distinguish the current journey from other journeys offered by the SP experiment. Indeed, there would be a case for developing different versions of the experiment, some with the reference alternative, and some without.

The analysis should then proceed to estimate VTTS by size of time saving, for both gains and losses. Since there does not appear to be clear consensus on how small is „small', analysis should investigate various resolutions, from seconds to minutes. Ideally there should be examination of „relativities', e.g. the importance of a 5 minute time saving on a 10 minute journey as distinct from a 60 minute journey. Note that this would seem to push the experiment towards an interactive implementation (e.g. CAPI) rather than pen and paper.

The representation of gains and losses might consider various reference points, including those explored in the context of focus groups, as well as other references inherent within the SP design (e.g. where the choice set includes the „current' alternative) and model (i.e. acknowledging that discrete choice models are specified in terms of utility differences across alternatives). Moreover, the analyst should at all times be clear as to their interpretation of the reference point and, it follows, their interpretation of gain and loss.

The data should be further subjected to analysis using random parameters logit specifications, with the objective of revealing variability in preferences for different sizes of time saving, and the extent of „negative' valuations of journey time. It is our view that the latter phenomenon implies a mis-specification of the model, potentially leading to counter-intuitive results. In order to avoid such problems, we recommend that analysts should make use of flexible distributions with few a priori shape assumptions, and with estimated rather than imposed bounds.

Again following ITS/Bates (2003), analysis should investigate the prevalence of so-called „tapering' effects associated with particular thresholds of time saving and loss. It would be useful to re-assess the evidence for the shape of the implied indifference curves, given the apparent violation of convexity reported by the 2003 study[10]. Note that the nature of tapering effects may depend upon the context, and may differ for example between passenger and freight.

8.3 Options B2.2 and 2.3

8.3.1 Background

If the Department were to commission options B2.2 and/or B2.3 then this would imply an interest in reviewing, at a more fundamental level, the current position on STS. That position is to value all time savings - irrespective of magnitude or direction (i.e. gain or loss) - at the same rate, hence the notion of a „constant unit value' (CUV). As is widely acknowledged, the CUV assumption brings computational convenience; it meets „additivity' and „reversibility criteria' and, in the context of appraisal, allows unweighted summation of STS across travellers and journeys. One should not overlook the practical implications of dispensing with the CUV assumption; these would be very significant.

That notwithstanding, it is perfectly sensible, from time-to-time, to rehearse the rationale for CUV, and to consider whether any theory, logic and/or evidence has emerged that might give cause to review that position. The theoretical basis for CUV - in terms of microeconomics - has been subject to vigorous debate; a succinct summary of the arguments can be found in ITS/Bates (2003). In our view, this debate is mature; we are not aware of any recent contributions that would call for a reopening of the debate.

It is our view that the more significant challenges to CUV emanate from logical/evidential considerations, namely: 1) trends in travel behaviour/transport planning practice, and 2) potential analogies between STS and experimental evidence on small vs. large changes in money.

8.3.2 Trends in travel behaviour/transport planning practice

There is a case for considering the extent to which contemporary transport behaviour/practices might give us cause to review the CUV assumption. Some examples that could be relevant in this regard include the following:

---

[10] It is the assumption of differential constant VTTS that leads to violation of convexity as the different portions of the indifference map cannot be joined up.

Discontinuities in time budgets

There is some evidence indicating the presence of thresholds in time budgets, for example the maximum amount of time that a driver is able to spend on the road during any one shift, the need for an overnight stay on a business trip, or the minimum amount of time needed in order to carry out a given activity (i.e. de Serpa's „technical constraint'). In these contexts, STS could conceivably be influential (especially if amounting to minutes rather than seconds); they could make the difference between a driver/traveller being able or unable to fulfil a given commitment. The trend of contemporary transport behaviour/practice would seem to be towards more, not less, discontinuity. If we accept such arguments, then this would imply that, depending on the inherent discontinuities, a given quantity of time saving may be worth more to some journeys than to others. The challenge is to explicate those discontinuities in a systematic fashion, thereby allowing the aggregation of benefits from STS across different journeys (length, purpose, etc.).

Just-in-time logistics

This has been a significant trend in logistics planning for 50 years, although the implications of STS in this context remain unclear. Within the context of a rigidly planned system, STS may imply idle time. On the other hand, STS could make the difference between meeting/missing a given delivery slot and, by implication, affect the number of vehicles required to complete all deliveries.

8.3.3 Experimental evidence on small vs. large changes in money

A second, and arguably more pertinent, case for reviewing CUV arises from the accumulating literature on experimental economics. A paper of particular significance from this literature is Tversky & Kahneman's (1991), which establishes the concept of „loss aversion' in the context of riskless choice (noting that most SP analyses of VTTS have been conducted under conditions of certainty).

Tversky & Kahneman review evidence for the proposition that reference levels have a significant effect on preference, and propose a theory based upon the following 3 principles:

Reference dependence: i.e. the carriers of value are gains and losses defined relative to a reference point.

Tversky & Kahneman observe that conventional microeconomic consumer models do not specify preferences as a function of current assets, such that indifference curves are drawn without reference to current holdings.

Diminishing sensitivity: i.e. the marginal value of both gains and losses decreases with their size.

Diminishing sensitivity implies that sensitivity to a given difference on a dimension is smaller when the reference point is distant than when it is near. The paper distinguishes between this concept of diminishing sensitivity, which relates to the reference point, and the concept of diminishing marginal utility, which relates to the slope of the indifference curve. Tversky & Kahneman remark that, although similar, the two concepts are logically independent, since diminishing sensitivity does not imply that indifference curves are concave below the reference point. The paper also introduces the concept of sign independence, which implies that reference independence can be violated only when a change in reference turns a gain into a loss or vice versa. Moreover, it can be verified that sign dependence is equivalent to constant sensitivity.

Loss aversion and shifts in the reference point: i.e. the value function is steeper in the negative domain than in the positive domain.

The paper notes that although the basic intuition is that losses loom larger than gains, a shift in the reference point can turn gains into losses and vice versa, thereby giving rise to reversals of preference. In this way, the model is able to account for phenomena such as the endowment effect and status quo bias. Tversky & Kahneman acknowledge that although loss aversion reflects the large disparity between WTP and willingness-to-accept (WTA), other possible sources of discrepancy include income effects, strategic behaviour and the legitimacy of transactions. It is noted that economic theory fails to distinguish between losses and foregone gains.

The purpose of Options B2.2 and B2.3 would be to assess the validity of the points outlined above, and their implications for the analysis of VTTS.

8.3.4 Tasks involved in Option B2.2: Review of theory, evidence and practice

The review would encompass various strands relevant to our interest in STS, as follows.

Theory and evidence on STS

We are aware of a small but emerging experimental literature on time preferences, for example Kroll & Vogt (2008) and Ellingsen & Johannesson (2009). This first stage of review would collate evidence on STS from the transport literature (e.g. De Borger & Fosgerau, 2008), and identify consistencies/inconsistencies with the aforementioned experimental literature.

Theory and evidence on small money savings

In contrast to the literature on STS, there is a considerable body of knowledge on the valuation of, and attitudes towards, small money savings. There is naturally an interest in translating such insights to STS, although in so doing there are three important distinctions which analysts should acknowledge.

First, whilst money is readily storable, time is less so. A fundamental principle of VTTS is that time savings carry value only if they can be converted into productive use, and this will not always be possible. Some activities may require a minimum amount of time in order to enable consumption (de Serpa, 1971), and/or be restricted to particular schedules (Vickrey, 1969; Small, 1982) or diurnal constraints more generally. Taking these factors together, it becomes apparent that time is a less flexible commodity than money, with the scope for reaggregation or replanning considerably more restricted.

Second, there is a question as to whether reference dependent preferences observed in the money domain readily translate to the time domain; i.e. can a small time saving be (naively) converted to money units via an appropriate value of time, and be interpreted with recourse to the experimental literature on small money savings? Third, a point often overlooked is that the literature on small money savings is typically set in the context of money risk (a prominent case in point would be Kahneman & Tversky's (1979) Prospect Theory), whereas the analysis of STS is typically set in the context of certainty (that is to say, SP experiments usually present certain journey times). There is thus a question as to whether attitudes to risk observed in the money domain straightforwardly transfer to the time domain, e.g. does the predominant behaviour of risk aversion in the money domain also apply to the time domain?; does the degree of risk aversion observed in the money domain similarly translate to the time domain? The review would rationalise the literature on small money savings, and form a judgement upon the transferability of insights to STS.

Proxy goods/bads for STS

A third line of inquiry would be to investigate opportunities for drawing insight from proxy good or bad markets, for example queuing behaviour in banks/shops, web-based checking in for flights, WTP a premium subscription for quicker internet download, etc. These investigations would seek to contribute alternative perspectives on the proposition that STS carry value.

Inter-temporal preferences with respect to STS

Another potential dimension of the trade-off between time and money concerns inter-temporal preferences; i.e. decisions involving trade-offs among costs and benefits occurring at different times. In terms of time discounting using money outcomes, there is a reasonable evidence of „hyperbolic' discounting. People are given choices like:

A: £100 a year from now,

or

B: £150 two years from now,

Then they are asked:

A': £100 now ,

or

B': £150 a year from now.

People choose B over A, but A' over B'.

Within the context of inter-temporal preferences, a potential distinction between the commodities of money and time is their relative perishability. In this context, one might question the reasonableness or otherwise of the following choices:

Would you prefer:

A: 1 minute quicker commute every day,

or

B: 3 minutes quicker once a week?

Similarly:

A: 2 minute saving without [current constraint],

or

B: 5 minute time saving with [current constraint]?

A: 5 minute time saving on Wednesday,

or

B: 3 minute time loss on Thursday?

Moreover, the review would investigate potential synergies between the experimental literature on inter-temporal preferences and the transport literature on time constraints/budgets and activity analysis. The review would further consider the extent to which inter-temporal preferences for time might be influenced by the magnitude of time saving/loss and the prevalence of risk.

The possible relationship between STS and TTV

As has been alluded to above, in the experimental literature on small money gains/losses, an important contextual factor is the prevalence of risk. Indeed this is the context for Kahneman & Tversky's (1979) hugely influential paper on Prospect Theory, which combines analysis of reference points and choice under risk. There is resonance between that observation and the proposition, acquiring some support in transport policy circles, that STS and travel time variability (aka travel time risk) may in some sense be related. This strand of the review would seek to clarify, in formal terms, the nature of any relationship between STS and TTV, both at the level of the individual traveller, and at more aggregate levels, as might apply to conventional transport planning models. This aspect of the review might usefully distinguish between short run responses to STS (as captured by SP, perhaps) and the longer run and more uncertain planning horizons considered by the models. The outcome of this strand would be a clear statement on the distinction between the concepts of STS and TTV, and a framework around which future VTTS research might be best specified so as to dissect the two effects and/or take account of any confounding of effects, as appropriate/necessary.

The predictability of STS

Distinct from the previous point, an attendee at the stakeholder seminar postulated a possible relationship between STS and a further dimension of the travel time distribution, which could be worthy of investigation. This concerns the *predictability* of STS, and would address a research question of whether, for example, a predicted time saving of 10 minutes by 1000 travellers is “more reliable” than a predicted saving of 1 second by 600,000 travellers.

The ability of SP to capture STS

Finally, if the phenomena described above are deemed to be relevant, the review would assess both the ability of SP to capture these phenomena and whether this would call for changes to design practices.

8.3.5 Tasks involved in Option B2.3: Exploratory development of new methods

Whilst we would recommend the commissioning of Option B2.2 in advance of B2.3, it is possible that the review would conclude that there is a case for re-thinking methods of analysis for STS. Option B2.3 would therefore exploit insights gleaned in the course of the review to design and implement such methods. It is difficult in advance of such a review, and potentially prejudicial, to promote specific methods within this note. However, as an indication of the possibilities, we suggest that a promising line of enquiry would be to investigate the transferability of methods from experimental economics. This might involve the following research activities.

The design of a laboratory experiment, possibly involving collaboration with experts in experimental economics. The designing of this experiment would provoke a range of challenges, for example:

- developing an appropriate survey vehicle for representing time as distinct from cost
- introducing risk into the experiment, perhaps in terms of both money and time
- ensuring that insights drawn from the experiment are applicable to non-work journey time, as distinct from other elements of time

The implementation of this experiment would prompt further questions:

- whether subjects typically used by experimental economists (often university students) are appropriate for analysing the value of time
- whether sufficient data can be generated from the experiment to support econometric estimates of VTTS

Finally, and depending on the success of the laboratory experiment, **there might be a case for considering two questions**:

- to what extent does experimental evidence support or undermine existing evidence on STS from SP?
- if the evidence from SP is challenged, how might SP be improved and/or experimental techniques be rolled out from the lab to the field?

8.4 Conclusions

Our overall conclusion is that we think there is a reasonably strong case for attempting to bolster the evidence base in the way proposed in B2.1 above, since we think the Department's position is actually stronger than it often appears on this point. We think that Options B2.2 and 2.3 represent the way to go at a deeper level if there was an appetite for doing so. Perhaps one of the more pertinent strands of B2.2 would be the mooted link between STS and reliability, which has found favour in some circles. Whilst it would be helpful to seek clarification on such strands, B2.2 would however entail a risk of reopening the STS debate. From an academic point of view, perhaps through UKTRC, there would be interest in that. However, we have some reservations about the priority which should be given to B2.2/B2.3 as a policy driven piece of work.

9 RESEARCH TASKS THAT WOULD BE INVOLVED IN COMMISSIONING OPTION B3 (ON 'VTTS AND VTTR FOR FREIGHT & LOGISTICS')

9.1 Introduction

As described in Chapter 5, significant work on freight VTTS and VTTR is ongoing, and so there are therefore clear advantages in taking stock of that before embarking on an expensive data collection exercise. Furthermore, there are existing data sets that could be further analysed in the light of latest thinking. Obviously, there is the likelihood that the wrong questions were asked, but we are satisfied that much relevant data of use is available.

In what follows, we outline 5 possible approaches to Option B3, namely:

- Option B3.1: An international review
- Option B3.2: Stated Cost
- Option B3.3: Conventional Fixed Design Stated Preference
- Option B3.4: Adaptive Stated Preference
- Option B3.5: Network analysis to estimate any additional effects of journey time savings

A further approach, in principle at least, is RP; however this is not considered to be a viable option in practice, and is not therefore developed further within the context of the present chapter.

9.2 Option B3.1: an international review

In our opinion, the first step in any plan to develop values for, and understanding of, VTTS and VTTR, should be a review of current understanding and evidence. It would be a mistake to specify methodology too tightly at this stage - that follows in the next section. Data on UK rail freight VTTS and VTTR using a "factor cost" or "cost saving" approach has been collected this autumn by ORR, and this would presumably be available for further analysis. SP data from the 2003-04 BAH/ITS study for SRA is also available, and could be reanalysed. That data covers both road and rail, but only for traffic that could potentially switch to rail. Quite a large survey of general HGV and LGV traffic was undertaken as part of the recent DfT M6 Toll Road survey. Unfortunately this gave unfeasibly low VTTS (and VTTR) values, partly due to unwillingness to pay tolls (even hypothetically) in order to gain time savings. Nevertheless, some relativities (i.e. trade-offs between variables) might be carried across to help build the bigger picture.

Similarly, relativities might be carried across from studies abroad, particularly in neighbouring countries. Besides that, methodology is being developed abroad, particularly in the Netherlands, which currently uses SP-derived WTP values in appraisal as opposed to the factor cost (drivers' wages plus vehicle operating costs) values used in the UK. It is for these reasons that we feel that any review of current values and methodology should be international in nature, with at least one funded meeting of experts. As previously reported, some reviews have already been completed, as part of European governmental VTTS/VTTR studies, but the published papers are limited for our purposes. They have summarised past work too briefly, so that we are given average values at a point in time in a certain currency for particular sorts of traffic in a given country. It should not be our purpose to merely add to such lists but to understand better the findings of a few of the most useful studies, particularly in relation to any useful relativities that might be transferred to the UK case.

An international study would be a major piece of work, and the size and cost of such a study may be rather elastic. While it should certainly try to identify important gaps in our understanding, it should generally provide a fairly complete picture of how we should treat VTTS/VTTR and what values should be used in modelling and in valuing time changes.

9.3 Option B3.2: stated cost

By the "cost saving" or "factor cost" method we mean building up VTTS and VTTR from component parts, such as drivers' wages and vehicle operating costs. These two are actually the only two headings under which costs are considered in official WebTAG road freight VTTS methodology. The current method has no value for freight VTTR per se.

It will immediately be appreciated that this WebTAG methodology allows nothing for getting goods (the cargo) to destination more quickly. This can be expected generally to have a non-zero value for a variety of reasons. In practically all cases (excepting only empty movements and movements of waste) there will be a capital servicing cost of the goods being moved. This may be very small for most loads. For example, a typical load might be worth £10000, so even with a 20% marginal cost of capital (i.e. the interest rate on the most expensive borrowing – usually a corporate overdraft) this is only £2000 pa. For a one hour time saving, this gives (2000/(365x24)) = 22pence/hour. There may be additional costs due to deterioration in travel, or severe constraints on both departure and arrival times such that each hour saved has a particularly high value.

Stated Cost surveys seek to establish all components of factor cost by directly asking the transport managers concerned. For VTTR we are dealing with unscheduled and unpredictable journey time changes, so cost elements include: risk of stock-outs; management costs to notify receivers of the delay; labour costs in keeping loading/unloading staff waiting and possibly needing overtime payments; and various "cancellation" costs if arrival would be so late that other arrangements have to be made.

As outlined in Chapter 5, there have been some attempts to use this approach in the UK. In 2003, SRA commissioned a study using a method developed by BAH and, in 2010, ORR commissioned a follow-up which has recently reported (AECOM & ITS, 2010). The former concentrated on traffic that might move, or was moving, by rail and the latter on just traffic that was using rail. In both cases most requests for costs were met with non-monetary responses. Due to the variety in situations, it is difficult to phrase questions that are unambiguous but flexible enough to let the respondent give the relevant information they have. By making use of the „partial' information received, and using values "transferred" from similar firms, a matrix can be built up, which can look to the uninitiated like a straightforward spreadsheet. It is difficult to explain to users of the values that that is not the case.

Consider the following very simple example where we interview two firms, 1 and 2. The catch is that they each present us with only partial information, differing between the two firms, which must be supplemented by exogenous data or „guesswork'. Neither firm reports the tonnes moved per movement, so we estimate them as ET1 and ET2.

If firm 1 reports its costs as C1, we can estimate its cost per tonne (X1) as EX1 = C1/ET1, this being reported cost (C1) divided by estimated tonnes (ET1).

Contrastingly, firm 2 might not report cost but instead report cost per tonne, as X2. Our cost estimate is then EC2 = (ET2).(X2).

Our initial figures for cost per tonne are therefore EX1 = C1/ET1 for firm 1, and X2 for firm 2.

The point is that a computer spreadsheet showing COST, TONNES, and COST PER TONNE columns would normally be populated for each firm, with no indication of how to introduce changes. In fact, forceful clients may insist that the spreadsheet contains "left-to-right" logic connecting the 3 columns. If the 3 columns just mentioned were in spreadsheet columns A, B and C, in that order, then column C would be expected to be a formula saying C=B/A. If, as in our example here, that was not how the column entries were originally derived, then this could lead to misinterpretation. We can illustrate a case in point as follows.

Suppose that we subsequently receive excellent information that our tonnes information is incorrect, and that true tonnes (T) are half our previous estimates (ET).

i.e. T1=0.5ET1, T2=0.5ET2.

What should we do? It seems obvious that we update the TONNES column in our spreadsheet. After all, the existing entries were simply our original estimates (presumably on a consistent basis). We will say below what SHOULD happen, but for now we will assume that a "left-to-right" logic has been imposed on the spreadsheet, such that the entries in the COST PER TONNE column both double (EX=C/T instead of EX=C/ET). That is clearly incorrect, since firm 2 had told us their true cost per tonne, X2, which we had used with our tonnes estimate ET2 to get our cost estimate EC2.

What we should have done was to have persuaded the client that the true data in the spreadsheet (initially C1 and X2) should be respected, and the spreadsheet logic allowed to vary firm by firm (rather than be the same for all rows). Cells with actual data in might then be colour coded, to emphasise that all rows do not work the same.

Of course, C=B/A will be true for each row – it has to be. The point is that sometimes C&A will be known and B estimated, and sometimes C&B will be known and A estimated. Columns may depend on later columns, not just earlier columns. Changing A will not always change C, - it may change B instead.

The new cost per tonne estimates should be EX1 = C1/T1 = 2C1/ET1 for firm 1, and X2 for firm 2, i.e. only one changes.

Whilst acknowledging that some useful data can be unearthed by such surveys, we are extremely conscious of the poor quality of responses. These are due to: ambiguity; asking the wrong person; having no correction for exaggerated responses; obtaining responses in different forms that cannot be "spreadsheeted" (in the manner described above); the sparseness of numeric responses; and the complete lack of any statistical methodology to handle the responses and provide confidence intervals etc. Some questions along these lines might be asked as a lead in to an SP survey, but relying on them on to obtain official VTTS values cannot be recommended. Ideally, we would want a review to consider the recent ORR survey and try to derive lessons for the future, but surveys to tease out costs savings are unfortunately in their infancy.

NB. We have not seen any interview using this method that could be described as "complete". Sometimes no numbers are obtained that were not already available on the web. Sometimes interviewers would promise numbers later, but then involve lawyers or otherwise delay beyond project end. A survey of 50 large shippers of roadfreight might usefully complement the recent ORR survey, but the cost would appear prohibitive set alongside the increase in knowledge.

9.4 Option B3.3: conventional fixed design stated preference

This method forms the basis of the official Dutch VTTS and VTTR values, and was used in the 1999 AHCG VTTS study for DfT. The method seeks to monetise respondent WTP/WTA for scheduled changes in journey time and the spread of journey time. The method is superior to Stated Cost (i.e. Option B3.2) since the offered choices have inbuilt trade-offs, and the method is fully supported by statistical theory. However, SP is not without its challenges. It is possible that respondents might misrepresent their relative valuation of money and time, leading to overestimates of VTTS; this is an intrinsic problem with SP, whether in freight or passenger contexts. Additionally, respondents facing penalties for late deliveries might respond correctly in the light of them but not consider if it would be cheaper to renegotiate the departure time. Any SP experiments that do not make provision for earlier departures are bound to overstate VTTS. Also, we have previously referred to possible unrealism inherent in presenting quicker road freight journeys with higher cost in non-toll situations. If we seek to overcome this by using a toll experiment then further problems could arise due to non-toll bias (as in the M6T freight study).

Regarding VTTR, overestimates will be found if the hypothetical choices force a significant chance of "late" arrivals, i.e. again preventing earlier departures. There has been a discussion as to whether VTTR should relate solely to the statistical concept of standard deviation of arrival time, or to the proportion of times arrival is "late". We need to know which assumptions apply in each case, which boils down to knowing something about the "shape" of the distribution.

Because conventional fixed design SP experiments find it difficult to allow respondents the choice between earlier departures and later arrivals, our view is that respondents are overly constrained and are likely to overstate their VTTS and VTTR in such experiments. A sample of at least 200 HGVs and 200 LGVs would ideally be needed. The great difficulty of finding that number of freight decision makers should not be overlooked. If the questions covered fall within the realm of driver decision making, then finding drivers should be possible. However, any decisions relating to differences in cost are likely to be controlled by management rather than individual drivers, though there may be rules for drivers to follow (such as when tolled facilities can and cannot be used). Discrete choice models are based on the assumption that (groups of) respondents are identical, and normal "averaging" rules do not apply. A sample made up from a small number each from a wide range of commodities, loaded and empty, travelling in HGVs and LGVs, on Own Account and Third Party (haulier) may only be able reliably to give one overall VTTS, with no way of reweighting it to be representative of any population of interest.

9.5 Option B3.4: adaptive stated preference

The characteristic of this method is that the later SP questions posed are based on the replies to earlier SP questions. ITS has been associated with this method for twenty years, using a particular methodology known as LASP. Face to face interviews are used, with the second half of the interview devoted to a computer based series of "screens", each containing 4 alternatives that have to be "rated". As far as we are aware, there are no other suppliers of ASP in the British market. LASP has been used in separate surveys of 40-50 respondents for TfL, HA, and SRA amongst others. LASP-derived monetary values are used both in the Great Britain Freight Model (GBFM) and the LEeds Freight Transport Model (LEFT). In those models the values are used for forecasting, and model calibration to past aggregate level outturn data makes it less likely that they will be overestimates. For appraisal purposes, as with conventional SP, there is a worry that respondents might spend money that in reality they did not have, or might try to inflate all the monetary estimates for strategic reasons. As with conventional SP therefore, it is to be expected that ASP will overstate VTTS and VTTR to some extent.

However, we would contend that the degree of any overstatement will be less for a well designed ASP experiment than for a well designed SP experiment. This is principally because rescheduling the journey is much easier in ASP. In real life, when there is a journey time change (such as might be due to a diversion for a planned bridge closure) freight operators have the option of changing either the departure time or the arrival time, or both. This is very difficult to incorporate into a conventional fixed design SP without sacrificing the rest of the experiment. In LASP, respondents are allowed to reject alternatives that are impossible for them, e.g. an arrival that is "too late". That alternative is then replaced (until accepted) and so no "screens" are lost, just a short delay in completing the experiment. Even if respondents do not reject alternatives outright, the adaptation procedure learns to offer alternatives that are not going to be poorly rated. The alternatives rated therefore tend to be more closely aligned with respondents' preferences than would be the case in a conventional fixed design SP.

The first part of each interview would permit the establishment of factor cost data, although some interview time would also be required to make the choice of the movement that is to be the basis of the ASP experiment and establish current data for that movement. As with conventional SP, averaging over disparate respondents is best avoided; rather it is sensible to aim for groups of about 5 similar respondents. This technique is therefore not one to take up in a hurry. Should a review identify a deficiency in understanding for half a dozen areas, e.g. specific commodities, then this method could be used to fill those gaps. Unlike conventional SP, it models each respondent separately – on the basis of the increased information content that adaptation makes possible for each survey. It is therefore perfectly proper to survey small numbers provided that it is understood that the results relate only to that small group. Combining to give values for bigger groups is possible, but requires complex weightings. Full statistical theory, giving confidence intervals and other statistics is applicable.

9.6 Option B3.5: network analysis to estimate any additional effects of journey time savings

This method uses a road network and determines the best locations for a network of depots to serve a set of customers. As point to point journey times change, due to things such as congestion, construction of new roads, etc., the optimal network will change. It has been observed that improved road point to point journey times over many years, particularly those when new motorways were regularly opening, led to much sparser networks. Travel had become cheaper in relation to depots, so companies chose more of the former and less of the latter. To some extent this was simply cashing in the time savings previously included in appraisals. It is possible to estimate the total costs for the firm in the two situations and hence derive the benefits to that firm of the journey time savings. This number could be compared to the valuation of those time savings using the official methodology. If the number were to be higher it would indicate that the current methodology was not capturing all the benefits of journey time savings, i.e. it was undervaluing them in these circumstances. Thus the thrust of B3.5 would be as an ex post rationalisation of travel time savings for the freight and logistics sector; its purpose would not be to deliver new estimates of VTTS/VTTR per se.

Some work in this vein has been done (e.g. Quarmby, 1989; Mackie & Tweddle, 1993), but none recently. The method is somewhat daunting, particularly regarding the estimation of depot costs. These vary enormously even for a given size because of the difference in land values in different locations. Also, depot costs vary not so much with throughput but with maximum design throughput, which will depend on temporal variability in demand and the firm's forecasts of demand many years hence. In our view, a further study in this area is timely and should achieve modest success.

The cost would depend on how big a study was required and whether any proprietary software would have to be purchased. There might also be costs in acquiring a suitable base data set. While it would be advantageous to work with recent data, this would raise serious questions of commercial confidentiality. In reality, customer locations change relatively slowly, in which case it might be possible to acquire several relatively old databases. There would still be a need for confidentiality but the data donators would be putting themselves at considerably less risk in the event of a breach. The next question would be whether the data was already coded up for a specific piece of software. With more than one dataset more than one piece of software might be required. Software producers are reluctant to allow access to their code, so the study might be restricted to the functionality built into the publicly available software. Such functionality may place constraints on network representation and, by implication, limits on how journey times might be changed.

In the light of the above uncertainties, it seems sensible to proceed first with a pilot project looking at just one dataset. Past UK studies appear to have used data from just one major supermarket chain, and we would advise trying for an alternative. Preferably, data for more than one firm might be combined. Once a potential dataset is identified, it must then be established what software it is (or easily could be) coded up for. It must then be established that that software can be manipulated in the ways required. One way forward would be for DfT to take the work this far internally, or on a call off consultancy basis. The pilot project should then conduct the analysis proper. Once this has been done for one firm, subsequent firms should be much easier and quicker, and the costs of further analysis should become reasonably clear. At the present time there is perhaps too much uncertainty regarding data and software availability to issue a contract on a fixed cost basis.

9.7 Conclusions

Whilst we acknowledge that the freight and logistics sector is difficult to analyse, we are doubtful whether current methods for VTTS (i.e. as detailed in WebTAG Unit 3.5.6) are adequate. Our overall conclusion is that we think there is good reason to commission research aimed at developing an update to/revision of VTTS and VTTR for freight and logistics.

Mindful of the analytical difficulties however, we believe that an appropriate starting point is the international review (Option B3.1). Significant work on freight VTTS and VTTR is ongoing, and there are clear advantages in taking stock of that evidence base before investing in new data collection exercises.

As regards fresh research, we believe that one area where the Department could do more to examine the relationship between transport and the economy is by studying the response of logistics networks to changes in travel time and reliability (Option B3.5). Work on this theme is now old and in need of updating. However, we are not convinced this strand of work will generate VTTS and VTTR which are usable directly in appraisal. Also the contractors with experience of logistics network models are not generally those who would undertake national value of time studies. We therefore recommend this is commissioned as a separate free-standing piece of work but within the same timescale as the main study.

At the same time as recommending B3.1 and B3.5, we are conscious that the Dutch VTTS/VTTR study is ongoing, which covers both passenger and freight. Although, as explained in Chapter 2, Dutch methods for valuing freight VTTS and VTTR differ in some respects from the UK's, the survey approach used by the Dutch essentially follows Option B3.3 (conventional fixed design SP). It would therefore be sensible (not least within the terms of B3.1) to explore this new evidence, to the extent that it could be used to inform the UK position.

10 A PROPOSED RESEARCH PROGRAMME

10.1 Introduction

Having outlined various research options in some detail, this chapter will endeavour to draw things together. We begin by summarising the full set of options, before collating our recommended options in the form of a research programme. We discuss the timelines of this programme, the resources required, and inherent risks.

10.2 Summary of options

10.2.1 Option A

Following from the earlier discussion of Chapter 6, Table 10.1 summarises the specific activities within Option A. Our strong recommendation is that Options A1 and A2 should form the core of any update to VTTS.

Table 10.1: Summary of Option A

Option	Description	Objective
A1	Repeat of 1999 car driver & pax survey	Updated non-working VTTS based on car mode
A1.1	Repeat of 1987 urban bus survey	Updated non-working VTTS based short distance PT mode
A1.2	Repeat of 1987 long distance coach & rail survey	Updated non-working VTTS based on long distance PT mode
A2	Update of car driver & pax survey to 2010 methods	Updated non-working VTTS based on car mode
A2.1	Update of 1987 urban bus survey to 2010 methods	Updated non-working VTTS based short distance PT mode
A2.2	Update of 1987 long distance coach & rail survey to 2010 methods	Updated non-working VTTS based on long distance PT mode

Option A1 would involve a repeat of the 1999 car driver and passenger survey, delivering updates to VTTS for the car mode, and allowing objective assessment of whether VTTS has changed over time (and whether the previous convention of uplifting VTTS has proved defensible). Provided the current assumption of transferability across modes is retained, A1 would allow straightforward review and revision of current recommended non-work values detailed in Table 2 of WebTAG 3.5.6.

By rolling out similar methods to other modes, Options A1.1 and A1.2 might yield insight on the validity of the transferability assumption (i.e. once individual-level characteristics of travellers have been accounted for, does VTTS shows variability by mode-specific features such as comfort?). However, if as is likely, there are found to be discrepancies across modes, that would leave a difficult practical question to be resolved (i.e. should any such discrepancy be attributed to mode-specific factors not explicitly modelled, or should different VTTS be adopted for different modes?). Remembering that AHCG (1999) focussed upon car, A1.1 and A1.2 would, if following the model of repeating previous national studies, need to look back further to MVA, ITS & TSU (1987) for a template to follow.

On balance, we feel that there is not a compelling argument for revisiting the transferability assumption, and we therefore consider Options A1.1 and A1.2 to be low priority.

Moving on to Option A2, this would similarly deliver updates to VTTS for car driver and passenger. However, whereas Option A1 sought to replicate 1987/1999 methods, Option A2 would update methods to present day standards and conventions (e.g. as documented in WebTAG guidance more generally). Options A2.1 and A2.2 would in principle employ the same methods to public transport, although our comments on A1.1 and A1.2 would apply analogously here.

10.2.2 Option B

With reference to Chapters 7-9, Option B encompasses three „areas of concern', namely:

- Option B1: reliability (VTTR)
- Option B2: small time savings (STS)
- Option B3: freight & logistics

We consider each of these in turn.

Option B1: Reliability

We believe there to be a strong case for exploiting insights from the ongoing Dutch study of VTTS and VTTR, particularly as VTTR can be a difficult area of research, and the Dutch approach might reasonably be considered to represent the state of the art. With reference to Table 10.2, we propose two alternative strategies for „piggybacking' the Dutch study. Option B1.1 would simply transfer estimates of VTTR from The Netherlands, whereas Option B1.2 would transfer methods from The Netherlands, but estimate VTTR specific to the UK. In principle, either approach would deliver updated estimates of VTTR for WebTAG 3.5.7.

Table 10.2: Summary of Option B1

Option	Description	Objective
B1.1	Transfer VTTR evidence from Dutch VTTS/VTTR study	Update non-work VTTR based on all modes
B1.2	Transfer VTTR methods from Dutch VTTS/VTTR study	Update non-work VTTR based on all modes
B1.3	Preparatory work	Prepare ground for transfer of Dutch evidence and/or methods
B1.4	Contingency work	Confirm validity of alternative approach to updating VTTR for car, should Dutch study fail

In order to inform the choice between B1.1 and B1.2, there may be value in first commissioning Option B1.3, which would sketch out the procedures involved in transferring methods and (especially) evidence from The Netherlands to the UK.

At this point in time, our recommendation would be for B1.2, since this would allow VTTS and VTTR to be analysed in combination, and would deliver UK-specific evidence. A drawback of B1.2 would be the need to delay the commencement of the entire study (i.e. VTTS and VTTR) work until the success of the Dutch approach had been confirmed. If on the other hand B1.1 were commissioned, then VTTS could proceed forthwith, with the Dutch evidence on VTTR combined downstream. Moreover, the attractions of B1.1 would be its cost and expedience.

Since the success of the Dutch study cannot be guaranteed, it would be sensible, at the same time as preparing for B1.1 or B1.2, to establish a contingency approach. This is the basis for Option B1.4, which would test the approach proposed by Fosgerau & Karlström, using existing data or fresh data from a pilot survey.

Option B2: Small time savings

As regards STS, Table 10.3 details three options, which proceed at different levels of ambition.

Table 10.3: Summary of Option B2

Option	Description	Objective
B2.1	Strengthen A1 and A2 methods to allow more detailed analysis of STS	Revisit the problem area encountered by the 1999 study, and strengthen methods using insights from ITS/Bates (2003)
B2.2	Review of theory, evidence & practice	Review Department's current position on STS, drawing upon theory, evidence and practical experience from transport and other relevant domains
B2.3	New methods for valuing STS	Blue skies research involving design and implementation of a laboratory experiment, drawing on expertise from experimental economics

The purpose of Option B2.1 would be to strengthen the methods (both focus group and SP) of Options A1 and A2, so as to allow more detailed analysis of STS, and thereby guard against the problems encountered by the 1999 study. We do not foresee that this option would provide enough evidence in itself to support a move away from CUV.

Option B2.2 would take a more fundamental look at STS, considering whether any theory, logic and/or evidence has emerged that might give cause to review the CUV assumption. One such issue, which has acquired some momentum in policy circles, is the proposition that STS and reliability are in some sense related. The review might therefore seek to clarify, in formal terms, the nature of any such relationship between STS and TTV, both at the level of the individual traveller, and at more aggregate levels, as might apply to conventional transport planning models. This option should deliver an authoritative review of STS, although the study would by its very definition be exploratory. Depending on the outcome of the review, it could strengthen or weaken CUV.

Option B2.3 would adopt a „blue skies' (i.e. not necessarily constrained by the fundamental principles which would seem to currently prevail) approach to STS. Responding to the recurrent critique that SP is inappropriate for the analysis of STS, this option would involve designing and implementing a new experimental vehicle, borrowing methods from the field of experimental economics. Whilst we believe that this option would entail exciting and valuable research for the medium term, we would not actively recommend this option as a component of any update study in the short term.

Option B3: Freight & logistics

Turning to the final „area of concern', we detail 5 options for the analysis of VTTS/VTTR for freight (Table 10.4).

Table 10.4: Summary of Option B3

Option	Description	Objective
B3.1	International review	Establish, as far as is practicable, authoritative evidence base on VTTS/VTTR by commodity flow
B3.2	Stated cost	Collect fresh survey evidence on VTTS/VTTR, to add to existing evidence base on stated cost
B3.3	Fixed design SP	Collect fresh survey evidence on VTTS/VTTR, to add to existing evidence base on fixed design SP
B3.4	Adaptive SP	Collect survey evidence on VTTS/VTTR, to add to existing evidence base on adaptive SP
B3.5	Network analysis	Analysis of changes in firm behaviour as result of road improvements, allowing validity check of current DfT approach to freight VTTS/VTTR

Before embarking upon any fresh data collection, which in the freight sector is notoriously difficult and expensive, we believe that there are clear advantages in taking stock of the evidence that already exists. Therefore we would strongly recommend Option B3.1. Although some reviews of VTTS/VTTR for freight have already been published, as part of European governmental VTTS/VTTR studies, these typically offer inadequate insight for our needs. This new review should not seek to repeat such work, but instead develop a deeper understanding of the evidence, in particular looking for relativities that could be transferred to the UK. Furthermore, this option could, where practicable and useful, further analyse existing data, for example ORR's data on rail time-related costs.

Jumping to Option B3.5, we suggest that any recommendations for update of VTTS/VTTR for freight and logistics could potentially be based on a range of sources, both methodological and evidential. With that in mind, there is a case, we believe, for investing in methods which could help to complete the evidential picture. It is against this context that B3.5 might be commissioned, complementing existing evidence from stated cost and SP. Following the line of Quarmby (1989) and Mackie & Tweddle (1991), a road network would be used to determine the best locations for a network of depots to serve a set of customers. As point-to-point journey times change, the optimal network can be expected to change, yielding insight on the manner by which freight and logistics operators respond to changes in VTTS and VTTR. Option B3.5 would not, in itself, deliver updates to VTTS/VTTR, but would serve as a validity check on the Department's current treatment of the freight and logistics sector.

The remaining options B3.2, B3.3 and B3.4 would collect fresh data collection using established methods, namely the methods of stated cost, fixed design SP and adaptive SP respectively. Whilst these options would bolster existing evidence bases, none of the data collection exercises would be trivial, and it is unlikely that any of these options would (in themselves, at least) deliver a convincing case for revision of recommended VTTS/VTTR for freight. Our inclination is therefore to prioritise Option B3.1, followed by B3.5.

10.3 Summary of option packages

The previous section has summarised the individual options. We now develop the discussion further by considering the manner in which the recommended options might combine to form a research programme appropriate for the Department's needs.

10.3.1 VTTS

With reference to Table 10.5, the first entry details Option A1. Since its methods are tried and tested, one would expect A1 to deliver key outputs at an appropriate level of precision and within appropriate timescale. If however any update to VTTS were reliant entirely upon A1, then it might stand accused of not making use of best analytical practice. For that reason, we recommend that A1 be combined with A2 (second entry of Table 10.5). By combining historical methods with current methods, A1+A2 would maintain consistency with 1987/1999 (depending on the modal focus), and mitigate many of the risks of A1 alone.

10.3.2 VTTR

Extending the scope of the programme to encompass VTTR, we would recommend that A1 and A2 be combined with Option B1.2. This package of options would comprise methods tried and tested in the UK and The Netherlands, together with current best practice prescribed by WebTAG and some emergent methods, delivering fresh UK evidence on VTTS and VTTR. The combination of VTTS and VTTR within a single survey would achieve cost economies relative to separate surveys for each, and allow investigation of potential duplication/confounding of VTTS/VTTR effects. Whilst we have expressed a recommendation for B1.2, we note that there exists a further credible option, namely B1.1, of directly transferring estimates of VTTR from The Netherlands. An attraction of this approach is that update of VTTS could proceed forthwith, with estimates of VTTR from The Netherlands combined later, albeit at the expense of not accounting for (explicitly at least) duplication/confounding of effects. Furthermore, B1.1 would be the cheaper option of the two.

In order to inform the choice between B1.1 and B1.2, there is a good case for commissioning Option B1.3, which would prepare the ground for the transfer of Dutch

VTTR evidence and/or methods to the UK. B1.3 could be commissioned as a standalone study.

Whilst the A1+A2+B1.2 package would, in our view, offer a reasonable chance of delivering key outputs at appropriate level of precision and within appropriate timescale, the key risk would be success of Dutch study, especially in relation to VTTR. For this reason, it would be prudent to establish a contingency in the form of Option B1.4

10.3.3 STS

Turning to STS, we would recommend one of two approaches, depending upon the position adopted by the Department. If the Department is broadly comfortable with the status quo, then there is good reason to commission Option B2.1. This should serve to bolster CUV, and defend against a repeat of the problems experienced in 1999. If, on the other hand, the Department wishes to retain an open mind on STS, then there is a case for Option B2.2. In particular, B2.2 might serve to confirm or refute the proposition, gaining momentum in some policy circles, that the phenomena of STS and reliability are related. Whilst clarification on such matters would be valuable, Option B2.2 could provoke a reopening of the debate on CUV more generally, and the Department should remain mindful of this possibility. Whereas Option B2.1 should, for reasons of economy and analytical convenience, be commissioned as an appendage to A1+A2+B1.2, Option B2.2 could be commissioned as a standalone.

10.3.4 Freight & logistics

Turning finally to freight and logistics, we are mindful of the difficulty and cost of sourcing data on VTTS and VTTR. Our recommended strategy, therefore, would be one of making best use of the available international evidence base, whilst selectively commissioning fresh analytical work. It is quite probable that any recommendations for update of VTTS/VTTR for freight and logistics would be based on a range of sources, both methodological and evidential. With that in mind, there is a case for investing in methods which could help to complete the evidential picture. It is against this context that we recommend Option B3.1, the international review of methods and evidence, followed by Option B3.5, the network analysis. With regards to the latter, we note that the analyses of Quarmby (1989) and Mackie & Tweddle (1991) are now somewhat dated; a further study would be timely and should achieve modest success.

Table 10.5: Summary of option packages

Option	Brief Description	Output	Initial assessment against objective	Risk	Project planning
A1	**Essentially a repeat of 1999 car survey** • No change to existing VTTS theories • Focus group + pilot + main survey + analysis • **Car only survey** • Non-working trip purpose • Pen & paper method	• Updated non-working VTTS based on car mode	• Would deliver key outputs, and maintain consistency with 1987/1999, but would not make use of current best practice methods as detailed in WebTAG	**Low-Medium** • Good chance of delivery, but issues concerning robustness of estimates, and potential for survey failures	**Should be core element of any VTTS study** • 12 months
A1+A2	**Update methods to present day standards and conventions** • No change to existing VTTS theories • Focus group + pilot + main survey + analysis • **Car only survey** • Non-working trip purpose • Pen & paper and computer-based surveys	• Updated non-working VTTS based on car mode	• Would deliver key outputs, and allow consistent comparison with 1987/1999, whilst also making use of current best practice methods as detailed in WebTAG	**Low** • Most sources of risk covered, as far as is possible	**Should be commissioned together with A1** • 18 months

A1+A2+B1.2	**Extend A1+A2 package to consider VTTR alongside VTTS using Dutch approach** • No change to existing VTTS theories • Focus group + pilot + main survey + analysis • **Car only survey** • Non-working trip purpose • Pen & paper and computer-based surveys	• Updated non-working VTTS and VTTR based on car mode	• Would deliver key outputs, and allow consistent comparison with 1987/1999, whilst also making use of current best practice VTTS and VTTR methods • Combines VTTS and VTTR in a cost effective fashion	**Low** • Most sources of risk covered, as far as is possible	**Significant economies by commissioning together with A1+A2** • 18 months
B1.3	**Preparatory work** • Short think piece preparing the ground for possible replication of Dutch VTTR study in the UK	• Description of rationale, and procedures to be followed	• Would articulate the basis for transferring valuations and/or methods from Dutch VTTR study, thereby facilitating B1.2	**Low**	**Could commission as a standalone study** • 3 months
B1.4	**Contingency work** • Pilot study of alternative approach to VTTR, in case Dutch study fails and roll out to UK not therefore practicable	• Estimates of VTTR based on pilot study, following Fosgerau & Karlström approach	• Whilst somewhat exploratory, this study should serve to demonstrate derivation of VTTR estimates from an alternative approach, thereby establishing a contingency to B1.2.	**Medium** • Exploratory, so outputs somewhat uncertain	**Could commission as a standalone study** • 3 months (6 months if fresh data collected)

B2.1	**Strengthen A1 and A2 methods to allow more detailed analysis of STS** • No change to existing VTTS/VTTR theories	• Analysis of STS using 1987/1999 methods, but strengthened in places to account for findings of ITS/Bates (2003) and provide for more extensive qualitative research	• Should deliver on objective, and thereby provide added reassurance that CUV assumption has been adequately defended	**Low** • Should deliver key outputs and bolster CUV	**Should be commissioned as add-on to A1+A2 or A1+A2+B1.2 packages** • No significant extension to A1+A2+B1.2 timelines
B2.2	**Review of theory, evidence & practice on STS** • Could have implications for CUV assumption	• Review of Department's current position on STS, drawing upon theory, evidence and practical experience from transport and other relevant domains	• Exploratory, and outputs therefore uncertain; could strengthen or weaken DfT's position on CUV	**Medium** • Wide ranging review, with implications for CUV unclear	**Could commission as a standalone study** • 6 months

B3.1	**International review of methods and evidence** • Might give cause to review means of disaggregating VTTS/VTTR for freight (e.g. by mode, by commodity) • Mainly desk top study, with some face-to-face interaction with international experts	• Improved evidence base on VTTS/VTTR by commodity flow, and better appreciation of methodological possibilities should fresh survey work be commissioned.	• Should meet objective, and thereby reassure stakeholders that any update of VTTS/VTTR would be based on a comprehensive and authoritative review of evidence	**Low** • A sensible way forward and should deliver key outputs	**Could commission as a standalone study** • 6 months
B3.5	**Network analysis of changes in firm behaviour as a result of road improvements** • Might give cause to review means of disaggregating VTTS/VTTR for freight (e.g. by mode, by commodity) • Mainly desk top study	• Validity check on current DfT approach to VTTS/VTTR for freight and logistics.	• Some significant risks, but should achieve modest success	**Medium-High** • Many risks, but should achieve some insights of value	**Could commission as a standalone study** • 6 months

Figure 10.1: Timed work programme for Option A1

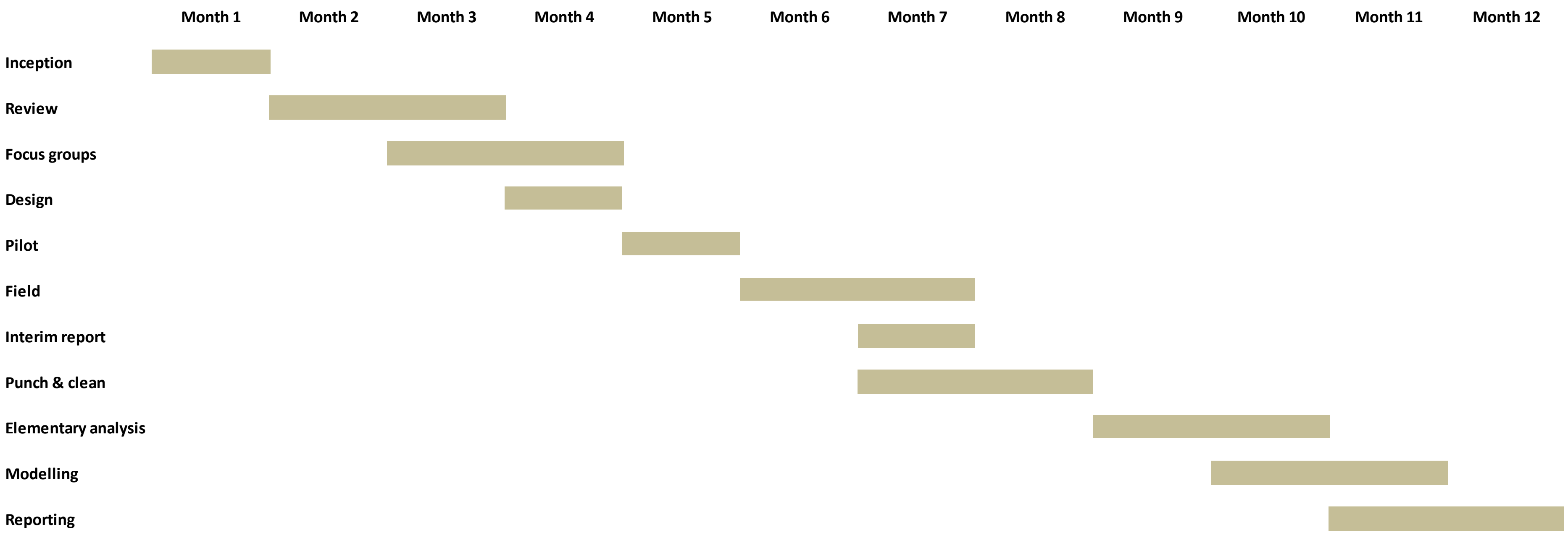
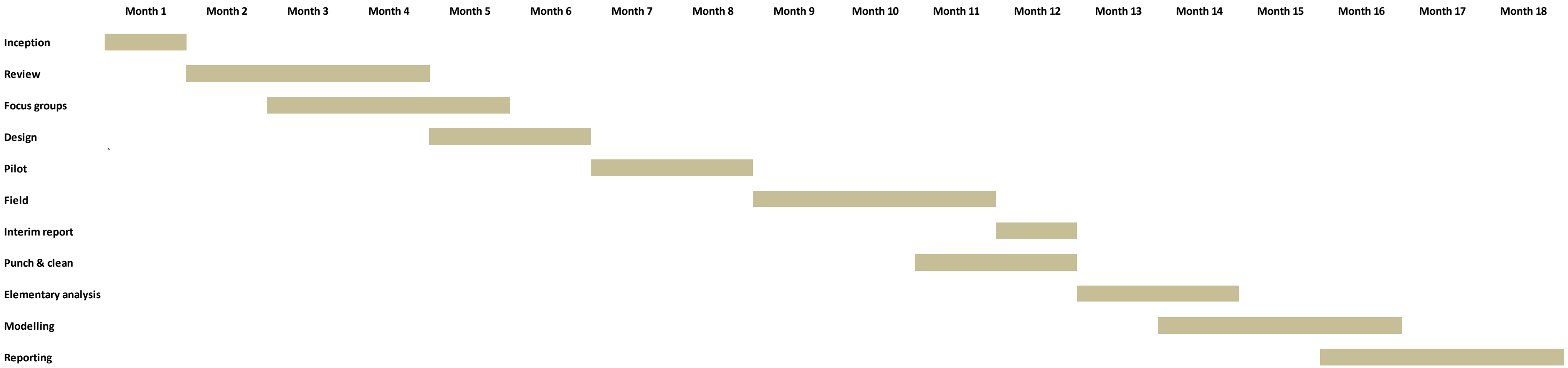


Figure 10.2: Timed work programme for Option A1+A2



10.4 Timed work programme

10.4.1 12 month work programme for A1

With reference to Figure 10.1, we present an illustrative timed work programme, as might apply to Option A1. As remarked earlier, we believe that A1 should form a core element of any update to VTTS.

We propose a 12 month work programme, on the basis that A1 would involve tried and tested methods. The programme makes provision for only a brief period of review, acknowledging that review outputs from Phase 1 would be at the disposal of the Phase 2 team.

Focus groups would proceed in a similar manner to the 1999 study, involving 6 focus groups of 8 respondents. On the basis that historical designs could be directly recycled, we legislate for only 1 month each for design and piloting work. Turning to the field study, this would again seek to mimic 1999, hopefully realising some economies in survey planning. Such economies would not however apply to actually distributing surveys and collecting returns, and it would therefore be important to allow adequate time for these activities (hence the provision of 2 months).

Turning to post-survey processing and analysis, this would also be largely a case of replicating 1999, and should therefore proceed reasonably quickly. Of course, some unforeseen issues may arise, for example concerning the cleaning of data, and it would therefore be sensible to leave some slack in the project plan to absorb such eventualities. We therefore allow 2 months for punching and cleaning, and a further 2 months for elementary analysis.

Option A1 should deliver recommendations on the update of standard valuations of non-working time per person, as documented in Table 2 of WebTAG Unit 3.5.6. We allow 2 months in the schedule for reporting to the Department.

10.4.2 18 month work programme for A1+A2

Whereas A1 involves tried and tested methods, A2 involves newer methods, and possibly some innovations. If commissioning A1+A2, it would be sensible, in our view, to extend the timed work programme to 18 months, allowing additional time for design, piloting and analysis. We illustrate such a work programme in Figure 10.2. A similar work programme should apply to A1+A2+B1.2.

10.4.3 Other work programmes

The A1, A2 and B1.2 options represent significant pieces of work, and will require careful planning. The remaining options are much smaller pieces of work, and whilst project planning will continue to be important, we are less inclined to prescribe specific work programmes to be followed.

Before concluding this discussion of the work programme, there would however be value in considering the chronology of the various commissions, so as to ensure coherent and efficient delivery of the requisite research outcomes.

Table 10.6: Phasing of options

Stage I: Review and contingency	Stage II: Survey and analysis
Option B1.3 VTTR preparatory work	
Option B1.4 VTTR contingency work	
Option B2.2 Review of STS	
Option B3.1 Review of freight VTTS/VTTR	
	Option A1+A2+B1.2 VTTS/VTTR survey
	Option B2.1 Strengthen A1+A2 for STS
	Option B3.5 Network analysis of freight VTTS/VTTR

With reference to Table 10.6, we would recommend broadly two stages of phasing.

Stage I would certainly include the review options (B2.2 and B3.1) and the VTTR preparatory work (B1.3), since these might directly inform the design of the subsequent survey and analysis options. Whilst not essential at this stage, it might also be appropriate to commission the VTTR contingency work (B1.4); this would avoid subsequent delays should the Dutch study fail, but at the expense of wasted resources should the Dutch study be successful.

Stage II would take insights from Stage I in the conduct of surveys and analysis. Option B2.1 would need to be commissioned in conjunction with A1+A2+B1.2. There is perhaps a question as to the most appropriate staging for Option B3.5. Since this work would be independent of the passenger VTTS/VTTR study, there would be no requirement to schedule this in parallel to A1+A2+B1.2. That said, it would be sensible to commission B3.5 subsequent to Stage I, thereby allowing relevant knowledge and evidence to be gleaned from B3.1.

10.5 Risk management

The risks inherent in the research programme as a whole are considerable, and the programme would require a well-conceived and well-implemented PPM plan in order to ensure delivery of outputs in a timely fashion and to the required standard. **In order to promote this objective, a key requirement, we believe, will be for the Department, as client, to establish clear and coherent lines of responsibility with respect to the various options that comprise the programme, and ensure continuity (for the life of the programme, at least) of staffing holding those responsibilities.**

The proposed research programme involves substantial data collection, and this introduces significant risks in project delivery, over and above the usual risks that any contract should legislate for. All bidders should therefore submit a Risk Management Plan as part of their PPM planning. This should anticipate risks, their consequences, their probability of occurrence, and establish appropriate mitigation strategies.

It is quite possible that contractor teams will be comprised of consultants from different organisations. Whilst this may allow teams to assemble the most appropriate range of skills and experience, the lead contractor will inevitably sacrifice some level of control over staffing and activities. This risk could be mitigated by establishing clear lines of communication and responsibility, reinforced by appropriate contractual arrangements.

It will be important to ensure that the timing and specification of outputs meet the Department's expectations and requirements. In order to support these objectives, the lead contractor should establish an open two-way communication with the Department's nominated project officer(s), documented through a regular reporting schedule.

Given the scale and complexity of the programme, engagement with the Department should be reinforced through the convening of a steering group. An important responsibility of the steering group will be to ensure compatibility between the discrete options that comprise the programme, especially where these are issued to different contractor organisations. Contractors(s) should report to the steering group on an intermittent basis, but particularly at key milestones, so as to review the strategic direction of their project(s).

In the conduct of any of the aforementioned options, a multitude of specific research risks will be encountered. Whilst not trivialising the significance of these risks, we leave it to bidders to anticipate these risks and develop appropriate risk management plans as part of their tenders.

10.6 Conclusions

Earlier sections of this report, not least our introductory comments of section 1.3, have highlighted various strategic decisions that will need to be made in finalising the design of this research programme. Whilst we will not seek to repeat these discussions, it is perhaps worthwhile summarising some key issues, especially to the extent that they might impinge on the planning of the programme.

1. **Should the update to VTTS simply be a repeat of the 1999 study?**

This would be a viable way forward, quick and cheap. However, it would be open to the accusation of not having made use of 2010 best practice methods.

2. **Should the update to VTTS be based solely on 2010 best practice methods?**

No, as this would not permit objective comparison against 1999. Our strong recommendation would be to combine 1999 and 2010 methods.

3. **Should the „areas of concern' be included within the scope of the VTTS update?**

In our view, VTTR should be an essential component of the programme. STS and freight are perhaps less critical, in that the approach to STS is governed largely by the Department's strategy (see questions 6 and 7 below), and freight could be developed as a self-standing interest (i.e. there is no direct interaction with passenger VTTS/VTTR).

4. **Should VTTS be surveyed together with VTTR?**

If VTTS and VTTR were to be taken from independent surveys then there is a risk that confounding between the two would not be adequately dealt with (although evidence from meta-analysis suggests that such confounding may be minor).

5. **Should VTTR evidence be transferred from The Netherlands?**

This would save a moderate amount of money, but may leave the study exposed to criticism concerning the sovereignty of evidence.

6. **Should CUV be reviewed?**

If the Department is broadly comfortable with the status quo, then there is a good case for commissioning B2.1. If on the other hand, there is a wish or pressure to undertake a more fundamental re-think, then B2.2 would be more appropriate, but the Department should remind mindful of the potential implications for policy.

7. **Should the possible relationship between CUV and VTTR be investigated?**

If answering „yes' then B2.2 should ideally be commissioned, but note comments under question 6.

8. **Should an update to VTTS include freight, or focus simply on passenger?**

There is a good case for reviewing the position on freight, but there would be no dependency on the passenger study; both interests could be progressed in parallel. Turning more specifically to freight, we believe that freight VTTS/VTTR would best be covered by commissioning an expert review (B3.1). Subject to the outcomes of that review, we believe that there is a good case for targeted research that seeks to complete the evidential picture, hence our recommendation for B3.5 on network analysis.

9. How can the programme be delivered to specification and on time?

This will be challenging. It will be important for the Department to put the right people and processes in place, and to ensure continuity at the „client end' throughout the life of the programme. Furthermore, it is imperative that the discrete work packages are tightly specified, and that one or more officers at the Department retain an overview of the coherence between these packages. It is clear from past experience that a shared understanding between Department and consultant of how exactly research results will feed into revised WebTAG and where the consultant's brief stops is vital.

11 CONCLUSIONS AND RECOMMENDATIONS

11.1 Background

The stimulus for this project was the Department's potential interest in commissioning a survey to update appraisal values for the Value of Travel Time Savings (VTTS). The last such survey was commissioned in 1994 (reporting in 1999), meaning that empirical evidence underpinning current official guidance on VTTS, as documented in WebTAG 3.5.6, is now up to 15 years old.

Mindful that any such update could be a significant undertaking, DfT commissioned the Institute for Transport Studies at the University of Leeds, together with John Bates Services and the Department of Transport at the Technical University of Denmark, to conduct a scoping study (referred to as „Phase 1') of the research activities that would be involved, thereby informing the planning of any subsequent implementation („Phase 2'). More specifically, the scoping team was tasked to deliver a range of options for the conduct of Phase 2, detailing an outline research specification for each such option.

The scoping study was conducted in broadly two stages, firstly review of history, methods and evidence, followed by consultation with the stakeholder community and scoping of the specific study options. With regards to the latter, we were required to consider: the scope of the research; the focus of the research effort; the structure of the Phase 2 study in terms of phasing; the data to be collected, with guidance on segmentation and sample sizes; and the methods to be employed in analysing the data.

The preceding chapters of this report have discussed these matters in some detail, and arising from these discussions, the scoping team is pleased to issue the following recommendations. We distinguish between recommendations which relate to broad strategy on the part of the Department, study design, the „areas of concern', and finally project planning.

11.2 Recommendations relating to broad strategy

A number of years have now passed since the last UK VTTS study, published in 1999, but based upon survey work conducted in 1994. During the intervening 15 years, many features of the UK transport sector have seen significant change, for example the provision of new travel options (e.g. toll roads, high speed rail, domestic air, LRT), the regulation of infrastructure and operations (e.g. the evolution of the rail industry post-privatisation), and changes in traveller behaviour (e.g. availability of travel information, changes in the way passengers use journey time). Throughout the same period, analytical method and practice have also witnessed significant developments, such that the evidence used to support the 1999 study - which forms the basis of current recommendations for VTTS as documented in WebTAG 3.5.6 - might, from the perspective of 2010, be viewed somewhat differently.

R1: There is a strong case for commissioning a revised national study updating VTTS.

11.3 Recommendations relating to the study design

As a framework for Phase 1, the Department instructed us to focus the scoping study around two broad options, namely Option A (which would review and recommend methods merely to update VTTS) and Option B (which would additionally address DfT's stated „areas of concern'). We believe that the Option A/Option B framework is well conceived and would be an appropriate way of structuring any update to VTTS.

R2: Any update study should be designed around a flexible and coherent structure comprising a core study of VTTS, with provision for additional modules dealing with „areas of concern' specified by the Department.

Within Option A, we considered two sub-options, namely Option A1 (this would, as far as is possible and sensible, repeat previous UK VTTS studies) and Option A2 (this would satisfy the same policy and practical requirements as Option A1, but exploit recent advances in methods). Again, we believe that the Option A1/Option A2 framework is an appropriate way of structuring any update to VTTS. However, we strongly recommend that A1 should be combined with A2. By combining historical methods with 2010 methods, A1+A2 would maintain consistency with the 1987/1999 studies, whilst adhering to current best practice, and mitigating many of the risks of A1 alone.

R3: The core study should comprise two elements. The first element should be a broadly faithful repeat of the previous national study in 1999. The second element should update analysis methods to modern day standards. These two elements of the core study should be commissioned together, and from a common contractor. The study should be designed so as to be capable of providing results for the VTTS as it varies with journey duration/length and journey purpose.

At the outset of the scoping study, the Department issued a long-list of 8 „areas of concern' which were to be considered within the context of Option B. Some of the „areas of concern' should, we feel, be adequately dealt with by a carefully conceived and executed analysis, and would not especially call for dedicated study options. Hence the provision for journey duration/length, group size and journey purpose under R3. We consider that two of the areas of concern, namely „further income segmentation' and „mode', are dictated more by policy convention than by analytical possibility. As regards income, convention is to calculate time savings for multi-modal appraisal using an „equity' value of time. As regards mode, official VTTS are based on car, and carried over to other modes. Whilst departure from these conventions would be analytically possible, this could have significant implications for analytical effort (and cost).

R4: On balance we recommend that the update study should focus upon the car mode, and retain the assumption of the 1999 study that, for appraisal purposes, VTTS is transferable across modes.

R5: The sampling strategy for the study will need to control carefully for income so that cross-sectional variation by income can be derived and any implications for income segmentation in modelling noted. Furthermore, comparison with 1999 values will strengthen the evidence base regarding the assumed income elasticity of the value of time. From the appraisal point of view, the choice between „equity' and income segmented values is a policy choice on which the study itself will not change the arguments rehearsed in previous work.

In our view, the three remaining „areas of concern'; „small time savings', „reliability', and „freight & logistics'; are pertinent to current policy trends, embody significant methodological and practical challenges, and provoke a range of unanswered questions concerning VTTS policy and practice. These should form the focus of Option B.

R6: We recommend that, beyond the core study, the update study should consider three „areas of concern', namely: VTTR, STS, and VTTS/VTTR for the freight and logistics sector.

11.4 Recommendations relating to the 'areas of concern'

11.4.1 VTTR

As we have already indicated, we believe that VTTR should occupy a key strand within any VTTS update study; indeed it should be the objective of the study to update *both* WebTAG 3.5.6 *and* WebTAG 3.5.7. There are, however, a number of alternative strategies which could be used to realise this objective, and these strategies could have different implications for costs, analytical convenience, project planning and policy. We would especially advocate the package of options A1+A2+B1.2. This package would comprise methods tried and tested in the UK and The Netherlands. Furthermore, the synthesis of VTTS and VTTR within a single survey would achieve cost economies relative to separate surveys for each, and allow investigation of potential duplication/confounding of VTTS/VTTR effects.

R7: We strongly advise that VTTR be surveyed in conjunction with VTTS, broadly following the approach used in the ongoing Dutch VTTS/VTTR study.

Whilst advocating Option B1.2, we are mindful that this introduces a dependency upon the success of the ongoing Dutch VTTS/VTTR study. It would therefore be sensible to establish a degree of insurance against the possible failure of the Dutch study, by means of Option B1.4.

R8: It would be prudent of the Department to additionally commission a pilot study of the „alternative' Fosgerau & Karlström approach to estimating

VTTR. This would establish a contingency, should the Dutch approach prove unsuccessful.

11.4.2 STS

We consider there to be two broad approaches to analysing STS. The first approach (Option B2.1) would involve collecting further empirical evidence on STS using existing methods. The second approach (Option B2.2) would review STS at a more fundamental level, involving the analysis of evidence from other domains (e.g. money as distinct from time), and potential interactions between STS and TTV. On balance, we would recommend B2.1 over B2.2, on the basis that B2.1 should serve to strengthen the current CUV position, whilst B2.2 is more speculative and may leave some challenges to CUV unresolved. B2.1 should, for reasons of economy and analytical convenience, be commissioned as an appendage to A1+A2+B1.2.

R9: On STS, we have reservations about whether even the best designed and conducted SP study will resolve the issues of (a) whether travellers behave according to the constant unit value assumption and (b) how this should influence the Department's appraisal conventions. We have some ideas regarding alternative research approaches but are not confident that these will resolve this longstanding problem area.

11.4.3 Freight & logistics

We are conscious of the ongoing contention regarding VTTS/VTTR for freight and logistics, and especially the proposition that recommended values, as detailed in WebTAG 3.5.6 and 3.5.7, fail to take proper account of changes in journey time/journey time variability on freight costs. Indeed, we share the basic aspiration to better understand the relationships between VTTS/VTTR and freight costs, but remain mindful of the recurrent challenges of sourcing relevant data, in sufficient volume, to shed light on those relationships. For this reason, we believe that a sensible way forward is to commission Option B3.1, which would undertake a detailed review of existing evidence, covering both UK and international sources. Conditional on the outcomes of that review, we also believe there to be a case for clarifying our understanding of relationships between VTTS/VTTR and freight costs, within the broader context of transport and the economy; such interests would be served by Option 3.5 on network analysis. We note that previous studies in that vein, by Quarmby (1989) and Mackie & Tweddle (1991), are now somewhat dated; a further study would be timely and should achieve modest success.

R10: With regards to freight, we would advocate a strategy of making best use of the available evidence base, whilst selectively commissioning fresh analytical work where this fills knowledge gaps. To these ends, we would recommend the commissioning of an international review of methods and evidence, followed by an analysis of the effects of VTTS/VTTR changes on freight network behaviour.

11.5 Recommendations relating to project planning

It would be sensible to devote careful consideration to the most appropriate phasing of the various options, so as to promote efficient delivery of the requisite research outcomes.

R11: We recommend broadly two stages of phasing for the update study. Stage I would include review work (especially on STS and freight) and preparatory work in relation to VTTR. Stage II (involving the core VTTS survey, plus ‚areas of concern') would take insights from Stage I in the conduct of surveys and analysis. There remains a question as to the most appropriate staging for the freight options, which would largely be independent of the passenger VTTS/VTTR survey.

A comprehensive VTTS update study would entail a major commitment on the part of the Department, demanding in both time and money.

R12: At an indicative level, we recommend that the Department plans for a research programme lasting 18-24 months in duration, and makes a budget provision of £500k-£750k.

The risks inherent in the research programme as a whole are considerable, and the programme would require a well-conceived and well-implemented PPM plan in order to ensure delivery of the requisite outputs in a timely fashion. Such a plan would, we believe, impose responsibilities not only on the consultant, but also on the Department.

R13: In order to promote delivery on time and to budget and specification, the Department should establish clear and coherent lines of responsibility at the client end, and ensure continuity (for the life of the programme, at least) of key staff holding those responsibilities.

ANNEX A: METHOD AND EVIDENCE FOR VTTS IN DENMARK

Evidential source of the national VTTS in Denmark

The Danish values of travel time saving (VTTS) are based on a series of projects undertaken for the Danish Ministry of Transport (MoT) and the Danish Social Science Research Council. An initial scoping study for the Danish MoT, led to a data collection project, which resulted in a comprehensive database available for the estimation of travel time values. The main content was a SP survey carried out via the internet supplemented by personal interviews for groups of respondents for which coverage via the internet was not sufficient.

Analysis of the data was carried out in two separate projects. The first was a research project financed by the Danish Social Science Research Council that developed the methodology for extracting travel time values from discrete choice data as well as delivering results of independent interest. This project resulted in the publication of a number of papers in scientific journals. It was very useful in establishing a solid base for the final project for the Danish MoT, starting one year later, aimed at obtaining travel time values for use in the Danish guidelines for transport project appraisal. The credibility obtained from the scientific foundation of the methodology turned out to be important for the eventual acceptance of results by the Danish Ministry of Finance, which has main responsibility for economic appraisal methodology in Denmark. The results of the applied project are described in (Fosgerau et al., 2007). Background information is presented in a series of notes.

In addition, the Danish national travel survey was used to reweight the sample to national values. The SP survey comprised four choice exercises:

- Experiment 1 (SP1): Abstract time-cost exercise examines trade-offs between in-vehicle travel time and cost;
- Experiment 2 (SP2): Disaggregated time components examines trading between hypothetical alternatives of the chosen mode and contains both in-vehicle and out-of-vehicle journey components (e.g. interchanges, access-egress, parking search).
- Experiment 3 (SP3): Alternative mode exercise considers time/cost trading for an alternative mode (i.e. not the chosen mode).
- Experiment 4 (SP4): Transfer price questions.

Data were collected for the following transport modes: car driver, car passenger, bus, metro, S-train and train. A parallel experiment for ferry passengers failed and the data from that were not used. Business trips were not included in the survey.

Analysis of data from experiment 3 was excluded from the project at its inception due to resource constraints. These data have later become useful for assessing various explanations for counterintuitive mode differences often found, i.e. that the VTTS for car drivers is larger than the VTTS for bus passengers, even after controlling for income and other background factors. The transfer pricing experiment failed and the data were not used. Thus the final VTTS was based on experiments 1 and 2.

In addition to the choice experiments, the survey collected background information concerning the respondents and the trip used as reference for the choice experiments.

There is a general discrepancy between on one side the level of detail and the effects included in the econometric models, and on the other side, the segmentation used in the final appraisal VTTS. In addition there is a gap between what may be learned econometrically and the need for definite appraisal values that can be applied. During the process and in the reporting of results, it was therefore made very clear where the econometric exercise ended, and which judgements that then were made in order to arrive at final VTTS to be applied in appraisals.

The approach chosen in the Danish study was to include everything in the econometric models of empirical relevance. The results were then presented to the steering group which could then proceed to decide how these results should be applied. The process is documented in the final report to make transparent what the decisions were and on which judgements the decisions were made.

Empirical evidence on passenger VTTS differences in Denmark

Models estimated

The central values of in-vehicle time were based on the first SP experiment. Models were formulated in log-willingness-to-pay space in order to achieve maximum control over the distribution of the VTTS. This follows the approach outlined in Fosgerau (2006, 2007). The data were segmented by mode, such that one VTTS distribution was obtained for each mode. All other segmentation was achieved by parameterising the VTTS distribution. There are a number of advantages to this approach:

- With present data, the model formulation yields considerable gains in model fit at low estimation cost.
- Reference-dependence is easily allowed for and can be tested.
- Errors are multiplicative such that the scale of the attributes of the alternatives does not affect choice probabilities in an undesired way.
- The model delivers the VTTS distribution directly, avoiding the need to derive the mean of a ratio of correlated random variables.
- It is possible to test the distributional assumption for the random component of VTTS by specifying generalised (flexible) distributions and testing against these.
- It is possible to provide a check of the identification of the VTTS distribution.

- Furthermore, the model allows the estimation a large number of significant parameters for the parameterisation of the VTTS.

The final models contained the following effects.

- Design variables
    - Significant evidence of loss aversion as predicted by prospect theory. Gains and losses are defined relative to the reference trip, which was a recent trip undertaken by the respondent. The SP design is pivoted around this trip.
    - The time and cost of the reference trip have significant effect on VTTS. The VTTS decreases with trip duration and increases with trip cost. This could be a selection effect.
    - The size of the travel time saving is significant. This is consistent with prospect theory. It was possible to accept that the VTTS stabilised for travel time differences above a certain threshold. It also seemed that this threshold increased with trip duration. The effect was interpreted as an editing effect, related to preprocessing of the choice situation by the respondent prior to evaluation, and was hence not carried forward to applied values. It is clear that much research remains to be done around this question. It is also clear that it makes no sense to include the effect in appraisal.
    - No significant anchoring effect. This is supports the use of stated choices to estimate values such as the VTTS.
    - Dummy for interview type was not significant.

- Trip related variables
    - Generally, there is no significant effect of travel purpose on VTTS. The models considered commuting to work, commuting to education, maintenance, and leisure. Presumably, the variation by travel purpose is captured by other traveller background variables.
    - Travellers with employer-paid trips have significantly higher time values than others.
    - Trip frequency has no significant effect on VTTS.
    - The effect of congestion on the VTTS is significantly positive for car drivers. An increase in the congested share of driving time of 25 percentage points increases VTTS with 27%. An increase of 50 percentage points increases VTTS with 60%.
    - Arrival time flexibility is significant for car drivers and bus passengers.
    - For car passengers, an accompanying child lowers the VTTS. This is probably not a causal effect.

- Socioeconomic variables
    - Log of net personal income has a significant, positive effect on the value of time. The effect of net household income (for couples) is also positive, but not always significant.

- A low income dummy is often significant, indicating a higher VTTS than can be explained by income itself.
- The age terms are only weakly significant
- In general, women have lower VTTS than men, the effect is significant only for car drivers.
- Family type is not significant.
- For car drivers we find that people in the Greater Copenhagen Area have significantly larger time values than the rest of the population.
- Home ownership has weakly significant positive effect on the VTTS.
- Most of the occupational variables have no significant effect. For car drivers, students have significantly lower VTTS than others.
- There is no effect of the interview type.

- Distribution of the VTTS
    - The test of a lognormal VTTS distribution is rejected for car drivers and bus passengers, but accepted for the other modes. The lognormal distribution was then applied for the other modes, while the more flexible distribution applied in the test (Fosgerau & Bierlaire, 2007) was applied for car drivers and bus passengers.
    - The range check indicates that the VTTS distribution is identified up to the 98 percent quantile in all modes. This is important, since lack of control of the VTTS distribution may lead to very large average values.

Calculation of mean values of VTTS for application

- Quadrant-specific effects related to loss aversion are omitted, leading to an average of WTP and WTA.
- The “employer-paid-trip” dummy is set to zero when calculating the mean VTTS.
- A constant VTTS is needed for application, but the VTTS increases markedly in the size of the travel time difference. The estimated thresholds were used as a guide in the absence of better arguments, emphasising that more research is called for on this issue. Based on various judgments, a range for the threshold of 10 to 20 minutes was recommended and a threshold of 10 minutes was chosen by the MoT on the basis of conservatism. Thus, the VTTS was computed for a time difference of 10 minutes, regardless of which time difference was presented in the interview.
- The mean VTTS is computed via simulation. The VTTS distribution is truncated at 1000 DKK per hour11, which mainly affected the car driver segment.
- The expected VTTS was then computed for each mode, reweighting the sample to be representative for the average kilometre travelled, which was felt

[11] About 120 GBP at the current exchange rate.

to be most appropriate for the applications of the VTTS. Other possibilities would have been the average trip or the average traveller. No segmentation by trip length was reported.

- The income effect was removed according to the estimated parameters.
- Finally, it was decided to use in application a single VTTS that did not distinguish between modes. The result was a unit VTTS of 67 DKK per hour of in-vehicle time.12 This is about two thirds of the net after tax hourly wage rate in the sample.

Relative values of travel time

The second experiment was used to determine VTTS for types of time other than in-vehicle time. These values were determined as factors to multiply the base in-vehicle VTTS. There were three variants of the second experiment. Car, single public transport mode, and multiple public transport.

For car drivers/passengers the considered travel time components are

1. free flow driving time,
2. additional driving time due to congestion,
3. access/egress walk time,
4. time spent searching for a parking space.

If any of the three latter components was not relevant for the subject, the variable was dropped from the design. There entailed a flaw in the SP design, which caused problems with dominant alternatives. All alternatives have the same transport mode as on the reference trip (car driver or passenger). Alternatives were created as pivots around the reference trip.

For public transport users, travel time is decomposed into

1. access/egress time (other modes than public transport, i.e, walking, cycling, taxi, etc.,
2. in-vehicle time (separate component for each transport mode),
3. headway of the first used mode,
4. number of interchanges between modes and associated waiting time (interchange waiting time).

In the single mode public transport experiment interchanges are between two vehicles of the same type, e.g. two busses. All alternatives used a single transport mode, which was the mode from SP1, namely the main mode of the reference trip.

For multiple-mode public transport users, the experiment was divided into two parts. In the first half of the experiment, variations in time were examined holding the number of interchanges constant. It contains alternatives with the two public transport modes used for the journey to allow valuations to be obtained for in-vehicle times for

---

[12] About 8 GBP at the current exchange rate.

each mode and out-of-vehicle time. The second part of the experiment allows the number of interchanges to vary by reducing one of the presented alternatives to use a single mode, allowing the value for between-mode interchanges to be valued.

Econometric models for these data allowed for

- Loss aversion in time components
- Nonlinear effect of headway
- Difference between short and long trips
- Individual specific random VTTS

The models used the logarithmic formulation of Fosgerau & Bierlaire (2009).

The final values that were decided for use in appraisal were the result of a discussion with the MoT concerning how to synthesise the results from the various models estimated.

Empirical evidence on freight VTTS in Denmark

Denmark has not carried out a study to assess the VTTS for freight.

Practice in CBA in Denmark - Passenger transport

Denmark uses a single base VTTS for appraisal that is common for all private transport across modes. This value applies to in-vehicle time. The VTTS is about two thirds of the net after tax hourly wage rate in the sample. Business travel is valued at the average wage rate multiplied by a factor to convert from factor to market prices.

The values that are applied to other time components are defined relative to the base VTTS as follows.

- Delay time for cars (i.e. time in addition to free-flow time) is valued at 1.5 times the base VTTS. This factor takes into account the uncertainty associated with travel time under congested conditions.

- For public transport, delay time, and wait time are valued at 2 times the base VTTS. Wait time includes half the headway of the first departure up to 12/2=6 minutes.

- Hidden waiting time is waiting time at the origin of trips using scheduled services, due to the headway. It is valued at 0.8 times the base VTTS. This value is applied to half the headway less 6 minutes.

- Interchange time is valued at 1.5 times the base VTTS. In addition there is a fixed penalty per interchange equal to 6 minutes at the base VTTS.

Practice in CBA in Denmark - Freight transport

The VTTS for the truck is based on cost accounting for truck and driver. The VTTS for the goods is based on the HEATCO project, with one unit value per ton-hour for road freight and another (lower) value for rail freight. The current values are 37.5 DKK per ton-hour for road and 15.4 DKK per hour for rail.13 Delays are valued at 1.4 times the value of free flow time.

Practice in CBA in Denmark - reliability

There is no official Danish practice for valuing travel time variability. A prestudy has been carried out (Fosgerau et al., 2008), building on the results in (Fosgerau & Karlström, 2009). The main issue in Denmark at the moment is the lack of models that are able to predict reliability for road transport. For rail transport, the current practice is to use the mean delay.

Future challenges

The Danish view is that the issue of the value of small travel time savings is important. It is seen as an artefact of the survey. So it is necessary to correct for the effect in order to obtain values that are relevant for CBA. The effect deserves more research, since the impact on the VTTS is large and it is recognised that a satisfactory solution to the problem does not exist at the moment. Having VTTS that differ by the size (and sign) of the travel time change does not makes sense in CBA.

The issue of travel time variability is high on the agenda in Denmark. So far a prestudy has been carried out (Fosgerau et al., 2008), which yielded a number of methodological insights. Most importantly, it showed how a value of travel time variability may be based on scheduling preferences (Fosgerau & Karlström, 2009). The prestudy concluded that it is possible to avoid the difficulties of requiring respondents to relate to probabilities by instead measuring scheduling preferences from which preferences for (against) travel time variability may be inferred. This approach is now being tested in the Norwegian study.

The prestudy also gave a recommendation for a value of travel time variability, based on a literature survey, which will be used until a Danish valuation study has been carried out. Finally, the prestudy gives recommendations for a future valuation study. At this moment, the traffic models that are in general use in Denmark are not able to predict travel time variability, and so it is often not possible to apply a value of travel time in specific analyses.

---

[13] About 4.4 GBP and 1.8 GBP per ton-hour at current exchange rates.

A value of time study for freight has been discussed for a few years. It is generally agreed that there is a need for such a study. On the other hand, it has not been given priority and, so far, other projects have been more urgent.

ANNEX B: METHOD AND EVIDENCE FOR VTTS IN NORWAY

Evidential source of the national VTTS in Norway

Norway carried out a VTTS study in 1997, which lead to a set of recommended VTTS in 1999.14 The 1997 Norwegian VTTS study addressed domestic passenger transport. Focus areas for the study were

- Distinction between short and long trips
- Time components
- Reliability
- Gain-loss asymmetry
- The effect of income

There is an ongoing new study that will address the VTTS as well as VTTR. At this moment, the study is coming out of the data collection phase and the analysis is beginning. Therefore, this note reports on the design of the new Norwegian study as well as the application of results from the previous study.

The new Norwegian study

The new VTTS study is a very ambitious study, covering many elements of VTTS and VTTR. The VTTS study covers the following modes:

- Long distance travel (+100 km): Car, Rail, Air, Bus, Ferry/Boat
- Short distance travel (-100 km): Car, Public Transport, Ferry
- Walk and Cycle as main mode of transport (not as access or egress modes)

The long distance study focuses on the valuation of:

- In-vehicle time
- Reliability of travel time
- Congestion (for car only)
- General “comfort” differences between modes

The short distance study focuses on the valuation of:

- In vehicle time
- Reliability of travel time

---

[14] TØI-rapport 459/1999.

- Congestion (for car only)
- Seat availability (public transport
- General “comfort” differences between modes

The Walk/ Cycle study focuses on the valuation of

- Walk/Cycle time
- Number of stops at intersections
- Provision of Cycle path
- provision of separate Walk and Cycle path
- Level of maintenance of Walk and Cycle paths

Each respondent in the first wave (VTTS study) get a maximum of 3 choice experiments, chosen with various routings from the following list:

- CE1: Within chosen mode, time and cost only (all respondents)
- CE2: Congestion, three attributes (car, long and short distance) or
- CE2: Reliability type (scheduling approach, scheduled modes, long distance) or
- CE2: Seat availability (scheduled mode, short distance)
- CE2: Reliability type (mean-variance approach)
- CE3: Within alternative mode, time and cost only
- CE3: Reliability type (scheduling approach)

This all respondents are exposed to CE1, one of the CE2 experiments and one of the CE3 experiments.

Users of slow modes are exposed to three different experiments:

- Between mode, slow vs alternative paid mode
- Separate walk/cycle path, no of stops or
- Walk/cycle path maintenance or

The within mode experiment is similar to that used in Denmark and Sweden and earlier in the UK and the Netherlands. There are two important modifications to the design. The first is that the range of bids has been extended relative to the Danish design such that the left tail of the VTTS distribution can be better identified. The second is that an off-reference choice has been added in order to be able to test some implications of prospect theory. In the design that formed the basis for the Norwegian design, all choice situations involved the both time and cost of the reference trip, in the same or in different alternatives. The off-reference choice is a choice situation in which the reference travel time, the reference travel cost or both do not appear in any alternative. Analysis of these data may lead to new insights concerning the issue of the value of small travel time savings. This is explained in De Borger & Fosgerau (2008).

The sequence of choices always ends by a WTP or a WTA type choice. The last choice is followed by a CV type question that aims to find the threshold at which the subject is indifferent between options.

The three-attribute experiments use the Bradley design and took the description in de Jong et al. (2007) as a starting point.

There are two types of reliability experiments. The first presents a range of five travel time outcomes associated with each alternative. In addition, the subject sees a trip cost. The second type of reliability experiment addresses reliability indirectly by measuring instead the scheduling preferences of subjects. Respondents are asked to choose between alternatives described by travel cost, travel time and minutes early/late at the destination. As shown by Fosgerau & Karlström (2009), scheduling preferences may be translated into a VTTR that takes into account the impact of the distribution of travel time. This is important since, under scheduling preferences, the value of standard deviation depends on the distribution of travel time. Another important advantage of this approach is that it is not required that subjects are able to digest the large amount of information that is present in a reliability experiment of the first type and subjects do not have to understand probabilities.

Empirical evidence on freight VTTS in Norway

Norway has not carried out a study to assess the VTTS for freight.

Practice in CBA in Norway - Passenger transport

The Norwegian VTTS for appraisal distinguish by

- Mode: car, train, air, and bus.
- Trip length: +/- 50 km
- Travel purpose: Business travel, commuting and other.

The distinction between short on long trips is generally not applied for road travel and an average value is used instead. This is presumably due to practical considerations.

A range of time components are often applied as indicated in the following table.

Table B1: Time components applied to official Norwegian CBA, by mode

Time component	Rail	Coast	Air	Road
In-vehicle time	x	x	x	x
Waiting time (actual)	x	x	x	x
Hidden waiting time	x			x
Access/egress	x			
Delay	x		x	
Interchange time	x		x	
Bicycle/walk			x	

Hidden waiting time relates to an interpretation of the effect of the headway of the first departure of a scheduled service. Even though passengers are not seen waiting at stops, they are thought to have scheduling costs, since departure times may not be optimal from the individual passenger's point of view.

Practice in CBA in Norway - Freight transport

There appears to be differences between the transport sectors concerning whether the time of the goods is valued. Vehicles, drivers etc. are valued at the value of alternative use.

Practice in CBA in Norway - reliability

Reliability is addressed by applying a standard multiple of the VTTS to average delays.

ANNEX C: METHOD AND EVIDENCE FOR VTTS IN THE NETHERLANDS

Evidential source of the national VTTS in the Netherlands

The national VTTS values for passenger trips (for all travel purposes) are based on the SP survey of 1997 (which closely followed the 1988 survey). The SP experiments contain only two attributes (time and cost), and refer to a trip actually made by the respondent (HCG, 1998). The realised numbers of returned interviews are 1,915 for commuting, 1,427 for business and 1,719 for other travel. The modes covered in this survey are car driver, train, bus and tram.

For freight transport, the national VTTS values are based on the SP surveys a carried out in 2003/2004. These experiments include as attributes for a typical transport of a firm (shipper or carrier): transport time, transport costs, unreliability (percentage not delivered on time), probability of damage and frequency. 194 firms were interviewed for road transport and 241 for the other modes (sea, rail, inland waterways and air transport).

Empirical evidence on passenger VTTS differences in The Netherlands

Information available in the new data

A new national SP study on the value of time and reliability values for both passenger travel and freight transport is underway (see Tseng et al., 2009). The study was commissioned by KiM, and is led by Significance. The survey includes experiments with travel time, cost, variation in time and arrival time. Also information is collected (for passenger travel) on observed travel time, cost and departure and arrival time, trip frequency, time constraints, journey distance, purpose, modes, congestion, use of information technology, household size and composition, household licence holding, car ownership, education level, gender and household net income (7 income bands). The sample size will be 5,200 respondents (with targets for car driver, public transport modes, air transport and inland navigation).

The data collection is planned to take place in October and November 2009. After that, discrete choice model will be estimated on the data. New VTTS are expected to become available in the summer of 2010.

Models estimated on the 1997 data

Both the national passenger VTTS studies of 1988 and 1997 used a specification where the time and the cost coefficient are interacted with various characteristics of the traveller and the trip:

$$U = \alpha T + \beta C + \sum \gamma i\ T\ Xi + \sum \delta j\ C\ Yj \qquad \text{(C1)}$$

where T is time and C is cost, with parameters α and β; the factors T Xi are adjustments of the time-parameter, based on personal or level-of-service variables and the factors C Yj are adjustments of the cost-parameter, based on personal or level-of-service variables.

Thus, the time parameter in a specific case can be expressed as:

$$\alpha + \sum \gamma i\ Xi \quad \text{(C2)}$$

whereas the cost parameter can be expressed as:

$$\beta + \sum \delta j\ Yj \quad \text{(C3)}$$

The VTTS is then given by:

$$(\alpha + \sum \gamma i\ Xi)/(\beta + \sum \delta j\ Yj) \quad \text{(C4)}$$

Using this general specification, separate models were estimated on the SP data for each of three travel purposes: commuting, business travel and other travel.

In the selected model for commuting (HCG, 1998, Gunn, 2001):

transport cost interacts with:

- income class (using dummies for 4 income bands; merging some of the original 8 that are available in the data), and

time interacts with:

- the type of household (e.g. presence of children)
- dummy for travellers journeying alone
- dummy for part-time workers
- age group (3 groups)
- gender
- amount of free time available (2 classes)
- mode (dummies for train and bus/tram)
- and motorway speed (4 classes).

The model for the employee's value of business time includes the following interaction variables:

transport cost interacts with:

- income class (using 4 bands), and

time interacts with:

- the type of household (e.g. presence of children)
- dummy for travellers journeying alone

- dummy for part-time workers
- age group (3 groups)
- gender
- amount of free time available (2 classes)
- mode (dummies for train and bus/tram)
- and motorway speed (3 classes).

In the selected model for other travel purposes:

transport cost interacts with:

- income class (using dummies for 4 income bands; merging some of the original 8 that are available in the data), and

time interacts with:

- the type of household (e.g. presence of children)
- dummy for travellers journeying alone
- dummies for part-time workers, for housewife and for pensioner
- age group (3 groups)
- gender
- amount of free time available (2 classes)
- dummies for sub-purpose is education and for sub-purpose is shopping/personal business
- mode (dummies for train and bus/tram)
- and motorway speed (3 classes).

The VTTS from this model will therefore be different for different values of these interaction variables. Most of these interaction coefficients were statistically significant (different from zero), though the differences between the estimated coefficients were not always statistically different from each other.

Subsequent analysis of the 1988 and 1997 data showed significant differences in VTTS by sign and size of the travel time change (e.g. Gunn, 2001, van de Kaa, 2005): larger time savings have larger unit VTTS (for „other' travel, changes of 5 minutes are not valued at all), and time gains are valued less than time losses. These research outcomes were not part of the VTTS study commissioned by the Ministry of Transport and have not been adopted in practice.

Empirical evidence on freight VTTS differences in The Netherlands

Information available in the new survey

In the new SP national survey for freight transport questions will also be asked about firm size, vehicle fleet and inventory size. For a selected typical transport of the shipper or carrier that is interviewed, the questionnaire asks about transport cost and transport time, departure and arrival time, time windows, commodity type, weight and value and packaging and containerisation. The sample will contain 520 firms, with targets by mode (road, sea, rail, inland waterways and air) and shipper versus carrier.

Models estimated on the 2003/2004 data

National freight VTTS surveys took place in 1992 and 2003/2004. As for passenger transport, discrete choice models were estimated on SP data. The 2003/2004 model produced different values for different commodities transported by road:

- Low value raw materials and semi-finished goods
- High value raw materials and semi-finished goods
- Final products, loss of value
- Final products, no loss of value
- Containers

There are also values for rail, inland waterways and sea transport (RAND Europe, 2004a), but these are not differentiated by commodity type, because there is considerably less variation in the commodity types transported using these modes, than in road transport.

Practice in CBA in The Netherlands - Passenger transport

Both the 1988 and the 1997 results for passengers led to a set of recommended values for cost-benefit analysis (CBA). The Ministry of Transport, Public Works and Water Management (e.g. AVV, 1998) issued a brochure containing the recommended values in the form of three tables:

- VTTS by income band (4 bands) and travel purpose (commuting, business and other)
- VTTS by income band and mode (car, train, bus/tram)
- VTTS by mode and purpose.

The car VTTS is for drivers, for car passengers the recommendation is to use 80% of the car driver value. Also car occupancy rates are provided (per purpose) for transport models that yield time gains per vehicle.

For commuting and other purposes, the VTTS are based on the SP models (expanded using the national travel survey). For business travel, the recommended values are based on the „Hensher' formula, where the employee's value comes from the SP among travellers, but which also contains an employer component, using productivity statistics.

Different values for different group sizes/compositions are not used (even though solo travellers were found to have a higher VTTS), because this additional complication (several transport models cannot produce this distinction in their output) was not considered worthwhile. All trip durations and distance classes receive the same unit value. This also goes for small and large time savings. This decision was based on the same arguments as discussed in the UK (actually The Netherlands is following UK practice on this). Nor is there a differentiation between VTTS by type of project (large, small, new infrastructure or pricing). An advantage of this uniformity is that different types of projects can be compared to each other.

These values were also published on the Ministry's website and included in the guidelines for CBA of transport projects (OEI guidelines). The Ministry has updated the values (for income/price changes) several times.

In practical forecasting, the national VTTS are of limited importance, because most transport models, including the national model LMS and the regional models NRM, have their own estimated coefficients for time and cost. The main motivation for the national VTTS surveys in The Netherlands was the use of the values in CBA.

Although the transport models used in most CBAs (LMS and NRM) can produce time gains by income band, in practical CBAs the distinction by income band was often not used, and only the values from the third table were implemented. The current tables on the Ministry's website are no longer by income band, but provide different values for different modes and travel purposes, as well as for different future years and four scenarios.

Practice in CBA in The Netherlands - Freight transport

The SP freight values from the 1992 and 2003-2004 studies have been used in several project evaluations. Often, the distinction between type of goods was not used in practice (due to lack of inputs that included this distinction), and the only distinction used was by mode. The recommended values that are currently on the website of the Ministry are by mode (road, rail, inland waterways, sea and air), future year and scenario.

Practice in CBA in The Netherlands - reliability

The recommended (preliminary) values of reliability are based on the outcomes of an expert workshop that took place on 25 October 2004. It was an initiative of the AVV Transport Research Centre of the Dutch Ministry of Transport, Public Works and

Water Management, and it was organised by RAND Europe. The aim of the workshop was to provide reasonable provisional values of reliability (VTTR) for a range of modes or mode-purpose combinations that can be included in the OEI-guidelines.

Within the workshop the focus was on valuation of the unexpected delays in travel time, preferably expressed as the standard deviation from the mean. While this definition has its limitations, using this definition would seem to lead to fewer ambiguities within the current cost-benefit framework.

A secondary aim of the workshop was to discuss ways to address the lack of sufficient studies with regard to VTTR for the various modes and purposes and explore options for international cooperation in addressing the research into evidence based monetary values for reliability.

Participants were invited, not only from The Netherlands, but also from the United Kingdom and Sweden, as reliability of travel time is an issue of growing concern and research in these countries.

The VTTR presented below is based on the opinions of the experts and the discussions during the workshop. These VTTR estimates are provisional values. To get evidence-based monetary values for reliability, a nationally representative SP study among car drivers, public transport users, carriers and shippers is now underway in The Netherlands.

The approach to measure the VTTR for passenger transport by car consists of the following steps:

- improved reliability of travel times is equal to reductions in travel time variability and thus to reductions in unexpected delays;
- The VTTR is defined as the value of a minute of standard deviation of the travel time distribution;
- In the workshop, the experts agreed that the VTTR is usually transformed into the Reliability Ratio:

  VTTR= RR * VTTS

  where:

  VTTR= value of one minute of standard deviation

  VTTS= value of one minute of average travel time

  RR= Reliability Ratio (=VTTR/VTTS).

  The basis for the agreement rested mostly on this being common practice: a convenient transformation since VTTS are usually available (and one wants to be consistent with these).

For passenger transport by car and public transport, the experts agreed on the following reliability ratios (based on the available international evidence, especially from the UK, The Netherlands and Sweden):

Table C1: Reliability Ratios for passenger transport by mode (purposes: commuting, business and other)

Mode	Reliability Ratio
Car	0.8
Train (interurban)	1.4
Bus, tram, metro (urban)	1.4

Source: RAND Europe, 2004b

The reliability ratios in Table D1 do not vary between travel purposes, but the fact that VTTS varies means that the same variation will be carried over to VTTR.

With regard to the application and dimensions of the VTTR for freight transport there was no consensus, nor even a majority position within the expert group. All experts agreed that much more research is necessary to establish VTTR for freight transport, in order to validate the obtained results of the Dutch study (RAND Europe, 2004a).

Meanwhile, SP values from RAND Europe (2004a) will be applied as the best available estimate, but these needed to be re-dimensioned into reliability ratios first. In RAND Europe (2004a), VTTR is measured through the scheduling method (that is relative to agreed delivery time). The research to convert these outcomes into reliability ratios is reported in de Jong et al. (2009). For freight transport by road, the recommended RR is 1.24. This value is preliminary and can be used until values from new empirical research in The Netherlands will become available (summer 2010).

The recommended provisional reliability ratios for passenger transport (especially for car users) are not based on research carried out in the Netherlands. In our opinion, such studies are required to get results that can replace the preliminary values, and this is the main motivation for the current value of time and reliability study. For freight transport, the large number of assumptions underlying the current (indicative) reliability ratios, was the major reason to undertake new empirical research in order to derive a more precise reliability ratio.

In practical CBAs sometimes just a markup on the travel time gains of 20% has been used to account for reliability gains.

ANNEX D: METHOD AND EVIDENCE FOR VTTS IN SWEDEN

Evidential source of the national VTTS in Sweden

The new Swedish Value of Time Study 2007-2009 includes VTTS for private travel (including commuting), but not for business travel (presumably because for business travel, the VTTS now used in cost-benefit analysis in Sweden is a pure cost-savings approach, that does not require surveys among the travellers). Travel time unreliability was not included in this new survey.

Two previous national freight VTTS surveys in Sweden that both used SP methods, are described in Widlert & Bradley (1992) and Henriksson & Persson (1999). There will be a separate new freight VTTS survey that will also include reliability. The pre-study on „Time and quality in freight transport' has started.

Empirical evidence on passenger VTTS differences in Sweden

Information available in the new data

The sample size of the 2007-2009 survey is as follows:

- Long distance bus and train: 1102 respondents (7966 observations)
- Regional distance bus and train: 1416 respondents (11529 observations)
- Car (1/3 long distance) 1449 (11593 observations).

The following variables are available for segmentation:

- Gender
- Net income
- Age
- Municipality
- Travelling alone/with others
- Household type
- Education level
- Employment
- Survey method (web/phone)
- Travel purpose
- Journey distance
- Journey time
- Travel Cost.

Models estimated on the new data

The models estimated in the new study on the SP data among travellers are not formulated in utility space, but in WTP space, as recommended by Fosgerau.

For private car trips, variables that have a significant impact on the VTTS are the following (from Börjesson & Algers, 2009):

- Higher incomes have a higher VTTS (elasticity of 0.45)
- Employed have a 29% higher VTTS
- Household with children under 12 years have a 44% higher VTTS
- Stockholm county has a 37% higher VTTS
- Observed travel distance or cost (elasticity of 0.20)
- Commuting has a 36% higher VTTS than other purposes
- The size of the travel time savings: larger savings have a higher unit VTTS (elasticity of 0.30).

For long-distance public transport, the following influences were found:

- Higher incomes have a higher VTTS (elasticity of 0.15)
- Employed have a 33% higher VTTS
- Observed travel cost (elasticity of 0.34)
- Observed travel time (elasticity of 0.09)
- The size of the travel time savings as proportion of trip time: relatively larger savings have a higher unit VTTS.

The VTTS results for local/regional public transport are as follows:

- Higher incomes have a higher VTTS
- Employed have a 21% higher VTTS
- Observed travel cost
- Observed travel time
- Commuting has a 22 % higher VTTS than other purposes.

Empirical evidence on freight VTTS differences in Sweden

Information available in the data

The INREGIA study (Henriksson & Persson, 1999) used SP experiments with transport time, transport costs, probability of delay and probability of damage. Six different commodity types were distinguished on the basis of the variables bulk/general, density and value density. The modes studied were road, rail and air transport.

Models estimated

Both the 1990-1992 and the 1999 surveys distinguished between values for road and for rail transport (though the latter study found the rail VTTS to be equal to 0). The 1999 study also produced a VTTS for air transport. No information on different SP values for different commodity types was found in the 1992 study. In the currently recommended VTTS for freight transport there is a differentiation by fourteen commodity types, which is based on the calculation of the average monetary value of the goods (per tonne, per shipment) for each commodity group, to represent the capital costs of the inventory in transit. This produced VTTS ranging from 0.05 SEK per tonne per hour (for the most bulky products) to 89.09 SEK per tonne per hour (for air freight).

Empirical evidence on the value of travel time reliability in Sweden

The 2007-2009 passenger VTTS study did not include the value of travel time reliability (VTTR). However, several empirical investigations on this topic have been carried out in Sweden. The most recent evidence concerns the VTTR for long-distance train trips (Börjesson & Eliasson, 2009). This work is based on SP surveys carried out in 2004 and 2007 on long-distance trains in Sweden. The VTTR is expressed as the expected delay time (the expected positive part of the difference between the actual and the scheduled travel time). The value of delay time turns out to depend on the delay risk level (expressed in the SP as the number of trains out of 10 that is late, or the frequency of delays per week). Lower risks have higher values of delay time.

Practice in CBA in Sweden - Passenger transport

Guidelines for carrying out a cost-benefit analysis (CBA) in transport have been laid down by ASEK, the official committee for CBA valuations. At the moment ASEK 4 (the latest update; see SIKA, 2008) is used, which includes VTTS based on the 1994 national Swedish value of time survey (see Algers et al., 1996). The new national VTTS survey (2007-2009) has not yet been finalised and is not yet included in the guidelines.

The VTTS for passenger transport in ASEK 4 distinguishes between:

Private trips:

- Regional and long-distance trips (where long distance is defined as over 100 km): the long-distance VTTS are usually higher
- In-vehicle time, headway, interchange time and delay time

- Headway is further distinguished by time bands (< 10 minutes, 11-30, 31-60, 61-120, >120 minutes): higher bands have lower unit VTTS
- Interchange time and delay time are further distinguished by mode (airplane or not).

Business trips:

- Mode (car, airplane, long-distance train, regional train and bus)
- In-vehicle time (the same value for all modes), headway (differs by mode), interchange time (the same value for all modes, except that not relevant for car) and delay time
- Headway is further distinguished by time bands (<60 minutes, 61-120, >120 minutes): higher bands have lower unit VTTS.

ASEK 3 of 2002 used the same distinctions, but in ASEK 4 the values have been adjusted for price and income changes. The biggest change from ASEK 3 to ASEK 4 was that the former used the „Hensher' formula for business trips whereas ASEK 4 now uses a pure costs savings approach. This implies that the current guidelines effectively assume that productivity during the trip equals zero and that all time savings in the long run are used as extra working time.

Practice in CBA in Sweden - Freight transport

The VTTS is freight transport in Sweden does not include effects of shorter travel times on vehicle operating costs (such as savings on staff time and on vehicles used), but only the effect that is related to the cargo itself (especially interest cost on the inventory in transit). Changes in the vehicle operating costs are included separately in the CBA.

ASEK 4 has different freight VTTS (in SEK per tonne per hour, or in SEK per lorry per hour) for fourteen different commodity types, as well as an overall value over all commodities. The freight transport models in Sweden also distinguish between commodity types.

Practice in CBA in Sweden - reliability

ASEK 4 includes a discussion whether the unreliability (variation) in journey times should be used or delay time, with references to the expert workshop on this in The Netherlands (RAND Europe, 2005) and HEATCO (2006). For private trips it decided to stick with delay time for bus, train and airplane (distinguishing between regional and long-distance trips), as mentioned in earlier sections.

For commuting trips by car a value for travel time uncertainty is recommended (90% of the in-vehicle time VTTS). For other private car trips, there is a value (50% higher than the normal in-vehicle time VTTS, following) for congested time.

The recommended values for business trips include a delay time for plane, train and bus (as mentioned earlier) and values for travel time uncertainty and for congested time for car trips.

For freight transport, ASEK 4 recommends a value of delay time which is twice the VTTS (in SEK per tonne-hour).

ANNEX E: META MODEL

For each value of time observation we have collected both continuous and categorical information, as apparent in the discussion above. The form of model used to explain variations in monetary values (V) takes a multiplicative (or constant elasticity) form:

$$V = \tau \prod_{i=1}^{n} X_i^{\alpha_i} e^{\sum_{j=1}^{p} \sum_{k=1}^{q-1} \beta_{jk} Z_{jk}} \qquad \text{(E1)}$$

where there are n continuous variables (X_i) and p categorical variables having q categories (Z_{jk}). We specify q-1 dummy variables for a categorical variable of q levels and their coefficient estimates are interpreted relative to the arbitrarily omitted level. The α_i coefficients are interpreted as elasticities, denoting the proportionate effect on the valuation after a proportionate change in X_i. The exponential of β_{jk} denotes the proportionate effect on the valuation of a level of a categorical variable relative to its omitted level.

A logarithmic transformation of equation E1 allows the estimation of the parameters by ordinary least squares (OLS) using equation E2.

$$\ln(V) = \ln(\tau) + \sum_{i=1}^{n} \alpha_i \cdot \ln(X_i) + \sum_{j=1}^{p} \sum_{k=1}^{q-1} \beta_{jk} Z_{jk} \qquad \text{(E2)}$$

Table E1: Meta-Model

	Coeff	T
Constant	-6.008	20.8
Attribute Specific		
Headway	-0.607	8.5
Late	1.134	7.3
Congested	0.316	2.6
DepShift	-0.282	3.2
ln_WalkTime	0.097	3.7
ln_WaitTime	0.072	1.6
Income		
ln_GDPindex	0.900	25.4
ln_GDPindex (EB)		

ln_GDPindex (Non-EB)		
Distance		
ln_Dist	0.156	7.1
+ ln_Dist_Car	0.047	2.5
+ ln_Dist_ Late_Inter	-0.239	3.9
+ ln_Dist _Other_Inter	-0.041	2.4
Purpose		
EB	0.746	13.8
PeakCommute	0.103	3.0
EB_Dep_Shift	0.384	1.9
Mode Used		
BusUser	-0.410	10.1
RailUser	0.360	7.0
+ RailUser_Interurban	0.242	3.6
AirUser	1.495	8.3
+ CarUser_Headway	0.187	2.1
+ CarUser_WalkWait	0.160	2.7
Mode (IVT) Valued		
RailValued	-0.157	-2.86
AirValued	0.584	3.14
BusValued_CarUser	0.162	2.68
Numeraire		
Fuel	0.082	1.2
Toll	-0.312	3.5
Data Type		
RP_IVT	0.144	1.6

RP_walk	0.266	1.9
RP_wait	0.268	1.5
SP Presentation		
Adaptive	0.283	2.0
www	0.611	2.5
Phone	0.351	3.6
Area		
LondonSouthEast	0.244	6.5
Other		
Mode_Choice	0.083	2.0
Ln_Comps	-0.072	4.0
Adjusted R^2	0.615	

REFERENCES

Advisory Committee on Trunk Road Assessment (ACTRA) (1978) Report of the Committee. HMSO, London.

Accent & Hague Consulting Group (AHCG) (1999) The Value of Travel Time on UK Roads. Prepared for Department for Transport.

Adviesdienst Verkeer en Vervoer (1998) Advies inzake reistijdwaardering van personen (Advice relating to the passenger value of time). Rijkswaterstaat, Ministry of Transport, Public works and Water Management.

AECOM & Institute for Transport Studies, University of Leeds (2010) Rail Freight User Values of Time & Reliability: Final Report to the ORR.

Algers, S., Lindqvist Dillén, J. & Widlert, S. (1996) The national Swedish value of time study. Paper presented at the PTRC value of time course and seminar, Easthampstead.

Atkins Consultants Ltd. (1997) Bus Reliability Study-Stated Preference Research, Great Britain.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J. & Bierlaire, M. (2008), Income and distance elasticities of values of travel time savings: New Swiss results, Transport Policy, 15(3), pp173-185.

Bates, J., Jones, P., Polak, J., & Cook, A. (2000) The Investigation of Punctuality and Reliability. Prepared for the Passenger Demand Forecasting Subscription Service.

Bates, J., Polak, J., Jones, P. & Cook, A. (2001) The valuation of reliability for personal travel. Transportation Research E, 37, pp191-229.

Bates, J. (2009) An agenda for research on reliability. Paper presented at the European Transport Conference.

Bates, J. & Whelan, G. (2001) Size and sign of time savings. Working Paper 561, Institute for Transport Studies, University of Leeds.

Batley, R., Wardman, M., Ibáñez, J.N., Shires, J. & Whelan, G. (2008) Revealed Preference Study to Assess Impact of Reliability on Passenger Rail Demand, Contract Number DFTRG/0027/2006, Deliverable 3: Report of Discrete Choice Analysis, Version 3.1, 24 October 2008.

Becker, G.S. (1965) A theory of the allocation of time. Economic Journal, 75 (299), pp493-517.

Beesley, M.E. (1965) The value of time spent travelling: some new evidence. Economica, 32, pp174-185.

Black, I. G. & Towriss, J. G. (1993), Demand effects of travel time reliability. Centre for Logistics and Transportation, Cranfield Institute of Technology, Great Britain.

Borger, B. de & Fosgerau, M. (2008) The trade-off between money and time: a test of the theory of reference-dependent preferences. Journal of Urban Economics, 64 (1), pp101-115.

Börjesson, M. & Algers, S. (2009) The Swedish value of time study. Presentation, WSP, Stockholm.

Börjesson, M. & Eliasson, J. (2009) Train passengers' valuation of travel time reliability. Centre for Transport Studies, Royal Institute of Technology and WSP Analysis & Strategy, Stockholm.

Bruzelius, N. (1979) The Value of Travel Time. Croom Helm, London.

Bruzelius, N. (2001) The valuation of logistics improvement of transport investments - a survey. SAMPLAN, SIKA, Stockholm.

Cantillo, V., Heydecker, B. & Ortuzar, J. de D. (2006), A discrete choice model incorporating thresholds for perception in attribute values, Transportation Research Part B, 40 (9), pp807-825.

Caussade, S., Ortuzar, J. de D., Rizzi, L.I. & Hensher, D.A. (2005), Assessing the influence of design dimensions on stated choice experiment estimates, Transportation Research B, 39, pp621-640.

Chintakayala, P.K., Hess, S., Rose, J.M. & Wardman, M.R. (2009), Effects of stated choice design dimensions on estimates. Paper presented at the inaugural International Choice Modelling Conference, Harrogate.

Copley, G., Murphy, P. & Pearce, D. (2002) Understanding and valuing journey time variability. Paper presented at the European Transport Conference.

Daly, A.J. (2009) The relationship of cost sensitivity and trip length. Paper presented at the European Transport Conference.

Daly, A. & Zachary, S. (1975) Commuters' values of time. Report T55, Local Government Operational Research Unit, Reading.

De Borger, B. & Fosgerau, M. (2008) The trade-off between money and travel time: a test of the theory of reference-dependent preferences, 64, pp101-115.

Ellingsen, T. & Johannesson, M. (2009) Time is not money. Journal of Economic Behavior and Organization, 72 (1), pp96-102.

Fosgerau, M. & Bierlaire, M. (2009) Discrete choice models with multiplicative error terms. Transportation Research Part B, 43 (5), pp494-505.

Fosgerau, M. & Bierlaire, M. (2007) A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B, 41 (7), pp784-794.

Fosgerau, M. (2007) Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A, 41(9), pp842-856.

Fosgerau, M. & Nielsen, S.F. (2007) Deconvoluting preferences and errors: {a} model for binomial panel data. Munich Personal RePEc Archive 2007 Sep 7;(\url{http://mpra.ub.uni-muenchen.de/3950/1/MPRA_paper_3950.pdf}).

Fosgerau, M. (2006) Investigating the distribution of the value of travel time savings. Transportation Research Part B, 40 (8), pp688-707.

Fosgerau, M. (2005) Unit income elasticity of the value of travel time savings. Oct 3, 2005.

Fosgerau, M., Hjorth, K. & Vincent Lyk-Jensen, S. (2007) The Danish Value of Time Study - Final Report. www.dtf.dk.

Fosgerau, M., Hjorth, K., Brems, C. & Fukuda, D. (2008) Travel time variability: definition and valuation. DTU Transport, Denmark, Report No.: 1:2008.

Fosgerau, M. & Karlström, A. (2009) The value of reliability. Transportation Research Part B, 44 (1), pp38-49.

Fowkes, A.S., Marks, P. & Nash, C.A. (1986) The value of business travel time savings. Working Paper 214, Institute for Transport Studies, University of Leeds.

Fowkes, A.S. & Whiteing, A.E. (2003) The evaluation of delays to freight on the Trunk Road Network, TRL Report TRL571.

Gunn, H.F. (2001) Spatial and temporal transferability of relationships between travel demand, trip cost and travel time. Transportation Research Part E, 37, pp163-189.

Hague Consulting Group (HCG) (1990) The Netherlands' Value of Time Study: Final Report. Report to Dienst Verkeerskunde, Rijkswaterstaat. The Hague, February 1990.

Hague Consulting Group (1998) The second Netherlands' Value of time study: final report. Report prepared for AVV, Hague Consulting group, The Hague.

HEATCO (2006) Proposal for harmonised guidelines. HEATCO Deliverable 5 (2nd revision), IER, Stuttgart.

Henriksson, C. & Persson, C. (1999) Tidsvärden och transportkvalitet - INREGIAS studie av tidsvärden och transportkvalitet för godstransporter. Report for the SAMGODS group. INREGIA, Stockholm.

Hensher, D. (1977) The Value of Business Travel Time. Pergamon Press, Oxford.

Hensher, D. A. (2001) Measurement of the valuation of travel time savings. Journal of Transport Economics and Policy, 35 (1), pp71-98.

Hensher, D.A. (2006), How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics, 21(6), pp861-878.

Hess, S., Daly, A. & Bates, J. (2005) Departure Time and Mode Choice: Analysis of Three Stated Preference Data Sets. Report to DfT.

Hess, S., Rose, J.M. & Hensher, D.A. (2008) Asymmetrical preference formation in willingness to pay estimates in discrete choice models. Transportation Research Part E, 44 (5), pp847-863.

Hess, S., Rose, J.M. & Polak, J.W. (2009), Non-trading, lexicographic and inconsistent behaviour. Transportation Research D, accepted for publication.

Hollander, Y. (2006) Direct versus indirect models for the effects of unreliability. Transportation Research Part A, 40 (9), pp699-711.

ITS, Imperial College & John Bates Services (2008) Multimodal Travel Time Variability, Final Report to the Department for Transport.

Jong G.C. de, Tseng, Y., Kouwenhoven, M., Verhoef, E. & Bates, J. (2007) The value of travel time and travel time reliability. Survey design. Final report for the Netherlands Ministry of Transport, Public Works and Water Management.

Jong, G.C. de (2008) Value of freight travel-time savings, revised and extended chapter for Handbooks in Transport, Volume 1: Handbook of Transport Modelling (Eds: D.A. Hensher & K.J. Button), Elsevier.

Jong, G.C. de, Kouwenhoven, M., Kroes, E.P., Rietveld, P. & Warffemius, P. (2009) Preliminary monetary values for the reliability of travel times in freight transport. European Journal of Transport and Infrastructure Research, 9 (2), pp83-99.

Jong, G. de, Tseng, Y., Kouwenhoven, M., Verhoef, E. & Bates, J. (2007) The value of travel time and travel time reliability. Survey design. Final report. For the Netherlands Ministry of Transport, Public Works and Water Management.

Kaa, E.J. van de (2005) Heuristic judgement, prospect theory and stated preference surveys aimed to elicit the value of travel time. Paper presented at the European Transport Conference.

Kroll, E.B. & Vogt, B. (2008) Loss aversion for time; an experimental investigation of time preferences. FEMM Working Paper 27, Otto-von-Guericke-University Magdeburg, Germany.

Kahneman, D. & Tversky, A. (1979) Prospect theory: an analysis of decision under risk. Econometrica, 47, pp263-291.

Lee, N. & Dalvi, M.Q. (1969) Variations in the value of time. Manchester School, 37 (3), pp64-78.

Mabit, S.L. & Fosgerau, M. (2006) Controlling for sample selection in the estimation of the value of travel time.

Mackie, P., Wardman, M., Fowkes, A.S., Whelan, G., Nellthorp, J. & Bates J.J. (2003). Values of travel time savings in the UK. Report to Department of Transport. Institute for Transport Studies, University of Leeds & John Bates Services, Leeds and Abingdon.

Mackie, P.J., Jara-Díaz, S. & Fowkes, A.S. (2001) The value of travel time savings in evaluation. Transport Policy, 37 (2-3), pp91-106.

Mackie, P.J. & Tweddle, G. (1993) Measuring the benefits gained by industry from road network improvements. Working Paper 391, Institute for Transport Studies, University of Leeds.

Mathematics Advisory Unit (MAU) (1970). MAU Note 179.

Metz, D. (2008) The myth of travel time saving. Transport Reviews, 28 (3), pp321-336.

MVA, ITS & TSU (1987) The Value of Travel Time Savings. Policy Journals.

MVA (2000) Evaluation of reliability and punctuality improvements on rail services. For the Shadow Strategic Rail Authority.

Orme, B. (1998) Sample Size Issues for Conjoint Analysis Studies, Sawtooth Software Technical Paper, http://www.sawtoothsoftware.com/technical-downloads.shtml#ssize.

Pinjari, A.R. & Bhat, C.R. (2006), On the nonlinearity of response to level of service variables in travel mode choice models. Paper presented at the Annual Meeting of the Transportation Research Board, Washington, D.C.

Quarmby, D.A. (1967) Choice of travel mode for the journey to work. Journal of Transport Economics and Policy, 1, pp273-314.

Quarmby, D.A. (1989). Developments in the retail market and their effect on freight distribution. Journal of Transport Economics and Policy, 23, pp75-88.

RAND Europe, SEO & Veldkamp/NIPO (2004a). Hoofdonderzoek naar de reistijdwaardering in het goederenvervoer (Main report of the study on the value of time in freight transport). Report TR-154-AVV for AVV, RAND Europe, Leiden.

RAND Europe (2004b) (Hamer, R.N., Jong, G.C. de & Kroes, E.P.) The value of reliability in transport; Provisional values for The Netherlands. Report TR-240-AVV for AVV, RAND Europe, Leiden.

Rose, J.M. & Bliemer, M.C.J. (2007), Stated preference experimental design strategies, Handbook of Transport Modelling, Hensher, D.A. & Button, K.J. (eds), Pergamon.

SDG (2003) The effect of road congestion on rail demand. For the Association of Train Operating Companies.

SDG (2005) Big delays. Prepared for the Passenger Demand Forecasting Subscription Service.

Serpa, A. de (1971) A theory of the economics of time. The Economic Journal 81, 828-846

SIKA (2008) Samhällsekonomiska principer och kalkylvärden för transportsektorn: ASEK 4. SIKA PM 2008: 3, SIKA, Östersund.

Small, K. (1982) The scheduling of consumer activities: work trips. American Economic Review, 72 (3), pp467-479.

Train, K. & Wilson, W. (2008) Estimation on stated preference experiments constructed from revealed preference choices. Transportation Research, Part B: Methodological, 40, pp191-203.

TRRL (1980) The Demand for Public Transport: Report of the International Collaborative Study of the Factors affecting Public Transport Patronage. Crowthorne, Berkshire.

TRL, Centre for Transport Studies University College London, TSU University of Oxford, ITS University of Leeds, Transport Studies Group University of Westminster (2004) Demand for Public Transport: A Practical Guide.

Tseng, Y., Verhoef, E., Jong, G.C. de & Kouwenhoven, M. & van der Hoorn, T. (2009) A pilot study into the perception of unreliability of travel times using in-depth interviews. Journal of Choice Modelling, 2(1), pp8-28.

Tversky, A. & Kahneman, D. (1991) Loss aversion in riskless choice: a reference-dependent model. Quarterly Journal of Economics, 106 (4), pp1039-1061.

Tweddle, G., Fowkes, A.S. & Nash, C. (1995) Impact of Channel Tunnel: a survey of Anglo-European unitized freight. Results of the Phase 1 interviews. Working Paper 474, Institute for Transport Studies, University of Leeds.

Vickrey, W.S. (1969) Congestion theory and transport investment. American Economic Review, 59 (2), pp251-260.

Vickrey, W.S. (1973) Pricing, metering, and efficiently using urban transportation facilities. Highway Research Record, 476, pp36-48.

Wardman, M., Toner, J.P & Whelan, G.A. (1997) Interactions between rail and car in the inter urban leisure travel market in Great Britain. Journal of Transport Economics and Policy, 31(2), pp163-181.

Wardman, M. (1991) Intertemporal variations in the value of time. Working Paper 566, Institute for Transport Studies, University of Leeds.

Wardman, M., Ibáñez, J.N., Tapley, N., & Vaughan, B. (2008): M6 Toll Study: Modelling of Passenger Choices. Prepared for the Department for Transport.

Wardman, M., Ibáñez, J.N. & Pagan, J. (2009) Variations in motorists' valuations of travel time with traffic conditions. Paper presented at Transportation Research Board Annual Conference. Washington DC.

Wardman, M., Ibáñez, J.N., & Hyman, G. (2008) The quality of a car journey: evidence from stated choice experiments. Paper presented at the European Transport Conference.

Warner, S.L. (1962) Stochastic Choice of Mode in Urban Travel: a Study in Binary Choice. Northwestern University Press, Evanston, Illinois.

Widlert, S. & Bradley, M. (1992). Preferences for freight services in Sweden. Paper presented at WCTR Conference, Lyon.

INSTITUTE FOR TRANSPORT STUDIES
DOCUMENT CONTROL INFORMATION

Title	Updating Appraisal Values for Travel Time Savings: Phase 1 Study
Authors	Institute for Transport Studies (ITS), University of Leeds, John Bates Services, & Department of Transport, Technical University of Denmark, under the aegis of the UK Transport Research Centre
Editor	Richard Batley
Version Number	1.3 (Final)
Date	20th June 2010
Distribution	Peter Mackie, John Bates
Availability	Pamela Chiang, Neil Shorten & Gavin Gaunt (DfT) and ITS project team
File	
Signature	