

Peer Review of Freight Value of Time and Value of Reliability

For National Highways and Department for Transport

<i>Version</i>	<i>Date</i>	<i>Author</i>	<i>Notes</i>
1.0	2/2/20224	Ian Williams	Draft for review by client
2.0	23/02/2024	Ian Williams	Major Expansion to Section 7, some other minor clarifications and additions, plus it addresses comments and queries of client on Draft 1.

Contents

Executive Summary

1	Introduction and context	7
2	Aspects to be examined in this peer review	8
2.1	Overview	8
2.2	Study design and sampling	9
2.3	Choice behaviour demonstrated in the survey	10
2.4	Modelling and enumeration methods	11
2.5	Magnitude of the VTT and VTTR results	12
2.6	Credibility of differences in valuations across vehicle type	13
2.7	Overlap between values derived and existing NF-VOCs	14
2.8	Impact of findings on freight modelling	14
2.9	Approach to adjusting the new values over time	14
2.10	Tests or further research?	15
3	Why representativity is important for freight transport analysis	16
3.1	Introduction	16
3.2	VTT and VTTR for HGV journeys to port	16
3.3	VTT and VTTR differentiation by commodity type, etc.	17
4	Selection of the sample	18
4.1	Sample design and potential for bias	18
4.2	Sample quota achievement	21
5	Design of experiments	22
5.1	Exclusion criteria	22
5.2	Transport cost relativities and BVTT levels in SP1	24
5.3	SP2 - reliability experiments design	30
6	Analysis of experiments	31
6.1	Issues for further consideration	31
6.2	Summary of issues	34
7	Usage in TAG	35
7.1	Original approach and its issues	35
7.2	The product related value of time and of reliability	36
7.3	Summary of implications for VTT/R estimation	40
7.4	Usage in TAG of the empirical results from the study	41

7.5	Recommendations for other TAG adjustments	44
8	Recommendations within future studies	46
8.1	Improving the segmentation within the sample	46
8.2	Sample size and survey design improvements	48
8.3	Experimental design improvements	49
8.4	Potential for revealed preference research	51
9	References	53

Tables

Table 1	Number of VAT and/or PAYE based enterprises by employment size bands for freight transport by road, UK 2023	20
Table 2	Number of responses, (Quota) for different vehicle types, destinations by shippers and carriers.....	21
Table 3	Sample sizes used for analysis, (Quota) for different vehicle types, destinations by shippers and carriers.	23
Table 4	For each segment: Observed mean journey cost per hour; for each bid rank - the average BVTT presented to respondents now illustrated as a proportion of this mean cost	25
Table 5	Cumulative percentage of number of times fast and expensive is chosen .	28
Table 6	For each segment, the % acceptance of BVTT when set: to 70% of its mean cost; or to the highest presented BVTT	29
Table 7	Estimated VTT by segment from early test of exclusion criterion (C1) and from finalised model	33
Table 8	Usability in TAG of estimated results by segment	42
Table 9	Relative scale of product related VTT and VTTR impacts, by model stage	42
Table 10	Percentage of journeys to / from intermodal terminals by type of HGV, GB.	47

FIGURES

Figure 1	Supply chains and logistic legs for consumer goods	39
----------	--	----

EXECUTIVE SUMMARY

This report provides a peer review of the findings from the recent research study, commissioned by National Highways and the Department for Transport, to examine road freight values of travel time (VTT), and values of travel time reliability (VTTR) in Great Britain.

The evidence used for this peer review is the detailed documentation within the Final Report (AECOM et al., 2023) that describes the methodology used and the results generated by the research completed by the study team of AECOM, Arup, Significance and ITS Leeds. The main stages of the study were: to assemble a suitable sample of road freight firms for both shippers and hauliers; to collect detailed information on these firms and their freight journeys made; to carry out a Stated Preference (SP) experiment; and to use choice models on the results of the experiment in order to estimate the freight VTT and VTTR for three classes of road goods vehicles. It was envisaged that these values could potentially be utilised to update the values currently in use within the Department's Transport Analysis Guidance (TAG).

The usage of SP experimental methods to estimate VTT and VTTR for freight transport, rather than just passenger transport, ensures that a very challenging, innovative and ambitious set of requirements faced the team charged with this research study. However, they have executed the vast majority of the required work to a good standard. Nevertheless, shortcomings have become apparent in a small number of methodological steps, which have compromised the applicability of the results.

The study successfully achieved the assembly of the required initial sample size of 602 firms to fully complete the set of SP experiments about a typical journey. This sample implied a response rate of less than 10% of the firms who had originally agreed to take part and it predominantly comprised larger firms. The sample was segmented into 12 detailed segments: between Shippers or Carriers (i.e. haulage firms); between three vehicle types (LGV, OGV1, OGV2) on the basis of the typical journeys of each firm; and between whether this journey was to port or not.

The quality control procedures applied to the initial set of results obtained from the SP experiments included 5 discriminating exclusion criteria checks. This caused the resulting sample available for analysis to reduce to 472, after one check was fully, and one was partly applied. This reduction ensured that only 2 of the 12 segments, namely the OGV1 shippers and the OGV1 carriers not to port, have achieved their quota of including at least 50 respondents, within this resulting analysis sample. This quota size was designed in order to facilitate effective parameter estimation explorations within the subsequent model analysis step. The other three exclusion criteria could not be applied as they would have reduced this sample available for analysis down to a totally unusable size.

The SP experiment for the VTT required each respondent to choose between a pair of travel options: one fast and expensive; the other slower and cheaper. This experiment

was repeated for 8 randomised option pairs, each of which had a different implicit Boundary Value of Travel Time (BVTT). The model analysis of these choices then identifies the VTT for that respondent. Although the set of BVTTs presented to a respondent were designed to vary around the reference journey time and cost reported as being typical for that respondent, the pattern of variations actually applied in practice appears to differ systematically between segments. The BVTT options were particularly low for the LGV segments, which led to many respondents being removed from the analysis for LGVs to ensure they would not corrupt the VTT model estimation procedures. This extreme pattern of BVTT option differences between segments suggests that many of the respondents in certain segments will have been presented with many BVTT option levels that differed substantially from their own expectations on mean journey costs. Perhaps this anomaly goes some way towards explaining the high proportion of respondents that had reported difficulty with the context or questions in the SP experiment.

For the model analysis tasks, the relatively small sample sizes that fell below quota for many of the segments, led to complications within the model estimation procedures. To resolve these issues, segments were combined so as to provide VTT and VTTR results that distinguished just the following 4 aggregate segments, rather than the originally planned 12 detailed segments.

- OGV shippers (i.e. OGV1 and OGV2 combined), further segmented by to port / not to port, though using a parameter that was not statistically significant.
- OGV carriers
- LGV carriers.

In the absence of a convincing estimated model for LGV shippers, it was assumed that LGV shippers should be allocated the same VTT and VTTR as OGV shippers.

The final study task was to transform the VTT and VTTR into a form that is suited to the requirements of TAG. The transformation method proposed summed the VTT of the carriers (the transport cost component) to the VTT of the shippers (the product related component). However the design of the experiment had a flaw that implied double counting of the transport cost element for a subset of the analysis sample for shippers, which in turn would overestimate the correct VTT for each segment. It would not be a straightforward analysis task to undo this double counting because this would require separating out subsets of shippers for individual modelling analysis. However the resulting sample sizes for such subsets would then be too small to facilitate successful model estimation.

The further analysis carried out within this peer review highlights: that the estimation of product related VTT and VTTR per tonne carried, is highly context specific; and that for any given vehicle type their values will vary widely and will be influenced by a wide range of inter-correlated explanatory variables. Accordingly, the estimation procedure would require a much larger sample than that which had been envisaged originally for this study, to ensure it could provide a sound model estimation of their values in a form that could successfully test the influence of all of these inter-correlated effects. This

would not be a simple modelling exercise and it is unlikely to be feasible to implement it through using the sample size available within the current study.

The carrier based VTTs for LGVs and OGVs estimated in the study are 4.1 and 3.5 times higher, respectively, than those currently within TAG. The TAG derivations are based solely on estimated driver costs plus vehicle operating costs per hour. Given the many caveats, including the lack of representativity of the samples and the apparently high level of lack of understanding of context among survey respondents, it appears prudent to not replace the current TAG approach and values until greater confidence can be reached about the solidity of the foundations of these new methods and of their empirical results.

On a more positive note, this peer review presents some suggestions for reanalysis of the sample results for OGV carriers, which might provide useful insights. It recommends examining for TAG whether the VTT to use for route choice modelling, mode choice modelling and for appraisal should all have identical values, or whether, say, the VTT used in appraisal should further include a wider range of longer term cost-saving adjustments within supply chains. It also recommends three other adjustments to the current TAG advice on aspects related to HGVs and LGVs.

With the benefit of hindsight from the current study, this peer review proposes a number of potential methodological improvements to future SP studies of VTT and VTTR. These improvements relate to adjusting the segmentation, increasing the sample size, refining the survey design, and to improving the design of the SP experiments. Finally, it outlines how revealed preference studies could be carried out in Great Britain as an alternative method for estimating freight VTT and VTTR.

1 Introduction and context

1.1.1 This report provides a peer review of the findings from the recent research study, commissioned by National Highways (NH) and the Department for Transport (DfT), that examined road freight values of travel time (VTT), and values of travel time reliability (VTTR) in Great Britain. These findings represent a significant addition to the evidence base, indicating the potential need for a change to the values embedded in the analysis underpinning business cases. As a result, there are various aspects on which the Department would like assurance prior to making changes to its Transport Analysis Guidance (TAG) to implement the findings of the new study.

1.1.2 The evidence used for this review of findings is the detailed documentation within the Final Report (FR – AECOM et al., 2023) which describes the methodology and the results of the research that has been carried out by the study team of AECOM, Arup, Significance and ITS Leeds.

1.1.3 The tasks specified for the peer review included the consideration of nine aspects of the research study and of its results. The conclusions from this peer review on these nine tasks are summarised in Section 2 below, together with references to where the evidence to support these conclusions has been assembled later within this review report.

1.1.4 The approach that has been taken in this review has focused particularly on the statistical methodology aspects underpinning the research, because any initial issues here would then permeate through to influence the conclusions and parameter values that could subsequently be derived from this study. Section 3 provides context on why the representativity of the sample is crucial for statistical analysis. Section 4 analyses the extent to which the sample of firms that were selected for the stated preference experiments for this freight value of time study are representative of road transport activity across Great Britain. Section 5 examines the set of contexts used in constructing the experiments to ensure that the research approach has provided a suitable and relevant cross-section of choices and to confirm that each individual's choice set is meaningful with respect to the experience of that specific participant. Section 6 focuses on the choice models that have been estimated from the results of the experiments. It assesses the model forms adopted and the decisions on which variables to exclude and why.

1.1.5 Having reviewed in detail the methodology that was utilised to generate the VTT and VTTR results, the implications for the potential uses of these results within TAG are considered in Section 7, which also includes recommendations for some other Guidance improvements relating to the modelling of road freight vehicles. Finally Section 8 summarises the main recommendations from this review of the findings from the VTT study with respect to potential improvements in the approach that could be considered for any future studies.

1.1.6 It is important from the outset to interpret and put into context any apparent lack of balance within the presentation throughout the rest of this review report between:

- its detailed focus on that subset of topics that represent areas for potential improvement in study design or implementation;
- versus the more summary discussion on the very many other areas where the research work described in FR is already to a good standard.

1.1.7 The desired output from this research study requires the design and execution of stated preference-based experiments to a level suitable for the successful estimation of highly segmented VTT and VTTRs for the road freight sector. This very challenging, innovative and ambitious set of requirements faced the team charged with this research study, and the vast majority of the required work has been carried out to a good standard. Accordingly, this review of FR and of the outcomes actually achieved, focuses particularly: on the lessons to be learned on how best to alleviate the various issues that have emerged along the way; as well as on those potential refinements to the existing approach to this task, which with the benefits of hindsight would have the potential to produce more cost-effective outcomes, if this VTT/VTTR research study were to be repeated and refined at some stage in the future.

2 Aspects to be examined in this peer review

2.1 OVERVIEW

2.1.1 The final report provides a good level of detail about the approach adopted, while the material generally is presented clearly in a well-structured fashion. The methods adopted for the estimation of the VTT and VTTR appear to have been implemented with care in most instances, though two errors in implementation appear to have arisen. Moreover, there were some aspects of the original study segmentation and design that were unhelpful. These together with the limited initial sample size and a strong likelihood of bias within it, have led to an inadequate sample size being available within the key model estimation stage. This then forced segments to be combined within the estimation to try to achieve an adequate sample size and it substantially reduced the quality of the resulting VTT and VTTR estimates.

2.1.2 Many of these methodological challenges arose because freight transport is inherently much more heterogeneous in nature than is passenger transport. Consequently freight analysis generates requirements for large samples that need to be initially selected and subsequently expanded with great care so as to enable them to be representative across the many diverse freight transport segments within the industrial structure of Great Britain. A large contingency margin would be needed within a study to enable changes in approach and in sample selection to be implemented in the course of that study in order to address the many and diverse challenges that are likely to emerge along the way.

2.1.3 The rest of this Section responds in turn to the nine specific aspects, copied across in italics below, that were specifically requested to be addressed within this peer review.

2.2 Study design and sampling

2.2.1 The study design and sampling, and whether the methods used are likely to have produced reliable and robust results. For instance, did the study utilise an appropriate and suitably representative sample of shippers and carriers, across (for example) business size, industry, region etc.

2.2.2 This peer review concludes that the sample of firms achieved was certainly not representative of the population of haulage firms across the UK. The sample predominantly covered just the largest haulage firms but these represent only 1% of the UK haulage firms as a whole¹, though due to having many vehicles these large haulage firms will carry substantially more than 1% of GB tonne kilometres. There also was a high 90% non-response rate within the original sampling, which runs the risk of ending up with a biased sample. The lack of representativity was also confirmed by the comparisons in FR Table 10 that present the sample versus the observed GB patterns of products carried, which indicated mismatches for a subset of product types. Likewise, the reported average costs per journey hour from the sample has indicated cost relativities between vehicle types of OGV1>LGV>OGV2². This does not match the expected sequence of OGV2>OGV1>LGV based on known average vehicle running costs by vehicle type from TAG.

2.2.3 However, successfully obtaining a good representative cross-section of firms in the sample is likely to be a very challenging task in practice. This is because the nature of this type of sophisticated stated preference data assembly exercise within a dispersed industry such as road transport, implies that those firms that have the time and the inclination to respond to the exercise are unlikely to be representative of the cross-section of firms across the road freight industry as a whole.

2.2.4 Obtaining a representative cross-section of firms was further complicated by requiring the journeys to ports by each individual type of road vehicle to be a specific focus of the study. This targeting is likely to have added bias to the sample, particularly because only a small and potentially atypical subset of all hauliers, other than those heavily using large articulated HGVs, are likely to generate regular movements to ports or airports.

2.2.5 The net result of the various challenges along the way was that the final sample of respondents available for the initial analysis was inadequate to achieve the required quotas for seven of the twelve segments that had been distinguished in

¹ See later Table 1 for IDBR data on the size distribution of road haulage firms. Data on average number of vehicles by size of haulage firm has not been located.

² LGV - Light Goods Vehicle up to 3.5 tonnes gross vehicle weight (gvw); OGV1 – Other Goods Vehicle over 3.5t gvw with 2 or 3 axles; OGV2 – Other Goods Vehicle with 4 or more axles (rigids) plus all articulated goods vehicles (artics).

the study design. This in turn led to considerable complications and limitations in the level of discrimination and precision that could be achieved within the subsequent model explorations for estimating the VTT and VTTR. Section 4.1 provides greater detail on the above topics.

2.3 Choice behaviour demonstrated in the survey

2.3.1 The choice behaviour demonstrated in the survey by respondents, in particular the non-switching behaviour demonstrated particularly by shippers and LGV respondents. What are the possible explanations for the patterns seen, and why has it impacted this study in particular? Crucially, it would be of value to understand the peer reviewer's view on whether this behaviour is likely to be a credible representation of stable, real-world choices from freight managers, or a byproduct of factors such as the survey design and the context of the study. On the role of survey design, were the Boundary Values of Time implicit in the study sufficient to generate the intended switching behaviour in the survey?

2.3.2 The transport cost values arising within this study gave rise to concern relating to two interdependent issues

- a) unexpected relativities of $OGV1 > LGV > OGV2$ between vehicle types for the typical costs per journey hour that were reported by respondents
- b) major differences between individual segments in the ratio of:
 - the Boundary Values of Travel Time (BVTT - which in effect is an implicit cost per hour) option sets introduced into their Stated Preference (SP) experiment;
 - compared to the typical cost per hour as reported by the respondents from that segment.

2.3.3 The first issue may well imply a major bias relative to the overall UK population of firms, in the characteristics of those sampled for some or all of the vehicle type segments. The need to include a full quota of firms that have regular shipments to ports/airports for OGV1s and LGVs may have contributed to such bias. Any such sample bias would impact on the validity of the results of the subsequent analysis unless some convincing method could be introduced to offset that bias. Unfortunately, the small sample sizes ultimately available for use in model estimation were insufficient to provide a set of significant external explanatory variables that could facilitate a differential expansion of the sample in order to mitigate bias.

2.3.4 The pattern of BVTT ranges that were presented to respondents within the SP experiments were radically different in their relative cost magnitudes across the set of different segments. In many cases the scale of these differences appears to have arisen due to obscure design mistakes / misinterpretations, rather than for productive reasons. These have resulted in modelling problems being generated for the shippers segments and for LGV carriers, through a loss of sample size due to the exclusion criterion (C5) needing to be applied. They may also have tended to bias upwards the estimated VTTs. Section 5.2 provides greater detail on the above topics.

2.4 Modelling and enumeration methods

2.4.1 The modelling and enumeration methods used by the research team, covering areas including (but not limited to) the model specification, the choice and application of exclusion criteria, and the use of CSRGT to make the sample nationally representative.

2.4.2 The selection of the five exclusion criteria that were examined during the study was appropriate and informative. However, only one criterion (C1) was fully applied, while another criterion (C5) was partially applied, to remove unsuitable respondents from the sample that was ultimately used for the model analysis.

2.4.3 Only a small minority of all respondents responded yes to the exclusion criterion (C4) question “D13 Did you find the description of the journeys clear?” The non-application of the three exclusion criteria (C2) to (C4) indicates that at least a quarter, and possibly over a half, of those ultimately included within the SP experimental analysis were respondents who had reservations regarding either the realism and/or their understanding of the context or questions that were used within their experiments.

2.4.4 The two exclusion criteria (C1) and (C5) that were actually applied caused reductions down to 472 in the effective sample size available for analysis. Consequently only 2 of the 12 segments of this study have achieved their required quota within the modelling analysis tasks. The further application of the remaining three exclusion criteria would have further reduced the available sample down to a size wholly inadequate for modelling, which is why they were not applied. There is some limited evidence that their non-removal may have biased upwards the VTT values, while it certainly will have reduced the robustness of the evidence underlying the parameter estimates. Section 5.1 provides greater detail on the above topic of exclusion criteria.

2.4.5 The inadequate sample size within the modelling stage, which arose for most of the initial 12 segments, has led to a number of simplifications and adjustments that were chosen to enable the model estimation to proceed and to mitigate impacts from not achieving quotas. The models ultimately estimated distinguished just the following four aggregate segments, rather than the originally planned 12 detailed segments.

- OGV shippers (i.e. OGV1 and OGV2 combined), further segmented by to port / not to port, though using a parameter that was not statistically significant.
- OGV carriers
- LGV carriers.

In the absence of a convincing estimated model for LGV shippers, it is assumed that LGV shippers should be allocated the same VTT and VTTR as OGV shippers.

2.4.6 For both shippers and carriers the initial estimates of VTTs for OGV1 were larger than for OGV2, which was contrary to expectations. Accordingly, the estimated VTT and VTTR were then forced in the model to adopt common values across the OGV1 and OGV2 segments.

2.4.7 It has not ultimately been feasible to identify the impact on VTT or VTTR of any external explanatory variables (e.g. commodity type or value, company size, stage in the supply chain, etc.). Because of this absence of external explanatory variables within the finalised models, the reweighting of SP results through use of CSRGT was necessarily limited in scope and so was unable to mitigate sample biases and to improve the national representativity of the estimated parameters. Section 6 provides greater detail on the above topic of limitations in the model formulations.

2.5 Magnitude of the VTT and VTTR results

2.5.1 The magnitude of the VTT and VTTR resulting from the study, and the likely driving factors for these. In particular, while the carrier values derived from the study appear to match expected magnitudes from international literature relatively closely, shipper and LGV values are substantially higher than suggested by previous literature. To what extent is this difference down to the fundamental choice behaviour discussed above, to UK-specific or recent freight industry factors (e.g. increased just-in-time deliveries and time-sensitive supply chains) or to other contextual factors around the time of the survey, including HGV driver shortages and higher-than-normal inflation? Are there any variables in the data, or model specifications not explored by the study team, that could shed light on this, or any adjustments that the DfT and National Highways could make to the resulting base valuations to account for these external factors? More generally, do we have a sufficient understanding of shipper decision-making to draw firm conclusions on the credibility of the results derived? Where possible, verification against industry cost data would provide additional assurance.

2.5.2 Unfortunately, various open questions remain about the reliability of the estimated values of the VTT and VTTR. There is evidence that the sample of respondents is far from being representative of the national population of hauliers. It was not feasible to adjust the results post-analysis to mitigate the impact of any such biases. The quota required to ensure that the model estimates for individual segments would be informative were only available within the finalised sample for two of the twelve segments. The resulting aggregation of segments in the models that needed to be adopted so as to mitigate the impact of the small samples, has reduced the level of discrimination within the results and has produced some unexpected relativities (e.g. VTT for OGV1 > OGV2). Accordingly, it would be difficult to decide the extent to which the higher than expected VOTs are a result of (a) real features particular to the situation in the UK in 2022, (b) rather than a side effect of the difficulties confronted in this study in the design and implementation of the SP methodology.

2.5.3 In particular, the approach adopted to distinguish between the VTT for a loaded and for an empty vehicle appears according to the documentation, to have included an error (see Section 7) causing it to substantially overestimate the loaded vehicle VTT. Due to the small overall size, 213, of the sample for shippers, there

does not appear to be scope to rectify this error by further excluding from this analysis sample the 125 shippers who outsource some of their transport.

2.5.4 A related important issue is that only a small minority of both shippers and carriers confirmed that the description of the journeys that was presented was actually clear to them. This reported lack of clarity suggests that doubts should remain on the ability of these respondents to make the required clear conceptual separation in practice within the experiments between the product cost related and the transport cost related aspects of the impacts of travel time changes. Section 7 provides greater detail on the above topic of the quality of the estimates of VTT and VTTR.

2.5.5 Drawing these findings together, suggests that the results unfortunately do not at present provide a conclusive quantified understanding of road freight decision-making.

2.6 Credibility of differences in valuations across vehicle type

2.6.1 Differences in valuations across vehicle type, and whether these appear credible. The LGV segment of the study produced particularly high values, potentially driven by a lack of switching in the stated preference games. Can the LGV results be considered reliable and robust for implementation in modelling and appraisal, and what LGV-specific factors should National Highways and the DfT consider, when considering the implementation of these values. Can the LGV sample be considered sufficient to produce robust results and would any small-sample regression methods help to resolve any issues in this area? Differences across OGV1 and OGV2 should also be considered, as well as the findings for trips to port/not to port, and how these could factor in modelling and appraisal guidance.

2.6.2 As already discussed across the previous sections, the average vehicle costs per hour initially reported by respondents for their own typical journeys did not have the expected relativities across vehicle types. This presumably indicates (in the absence of any other credible explanation) that this sample of respondents is biased and is not representative of the population of UK road freight firms as a whole. Accordingly, in the absence of any feasible post-processing options for adjusting and improving the representativity of the sample, it is perhaps not surprising that the resulting VTT relativities were not as expected across vehicle types. As discussed below in Section 7.2, an explicit representation of the stage within the supply chain of the journey considered in the SP, might have helped to understand better some of the apparently anomalous relationships between goods vehicle types in their estimated levels of VTT.

2.6.3 Due to the LGV sample sizes available for model analysis phase being below half of the required quota for every individual LGV segment, and due to inappropriate BVTT option sets for LGVs being used in the SP experiments, problems arose at this LGV modelling stage and accordingly the scope for in-depth modelling analysis was limited.

2.6.4 The small sample sizes may have been the cause behind why none of the models produced a statistically significant parameter for journeys to ports.

2.7 Overlap between values derived and existing NF-VOCs

2.7.1 The overlap between the values derived and existing Non-Fuel Vehicle Operating Costs (NF-VOCs) in TAG. In particular, the potential for the new study findings to replace the b1 parameter (time-related capital costs) for the relevant freight vehicle types. What implications would this have for modelling and appraisal?

2.7.2 For the reasons summarised in the previous sections, it would be premature to use this study's results as the foundation for wholesale changes to the existing TAG approach to Non-Fuel Vehicle Operating Costs. The study methodology that associates the carriers' results with the driver plus vehicle operating cost element and the shippers' results with the product related element of VTT, does not appear to have been implemented correctly in practice (see Section 7), so some of the current values produced by this study are questionable.

2.7.3 Prior to introducing the (future corrected) study results to replace the b1 parameter in TAG, it would be prudent to first carry out a direct confirmation that the study methodology could successfully generate the separation between vehicle related and product related VTTs that it aims for. This confirmation could be tested by adopting the more comprehensive revised SP approach that is proposed in Section 8.3 below.

2.8 Impact of findings on freight modelling

2.8.1 More generally, the likely impact of implemented findings on freight modelling, and transport modelling in general. What impact might the findings have for route-level freight traffic assignments, and result forecasts of traffic and congestion for all traffic. What implications do the findings hold for the appraisal of freight transport via non-road modes, such as rail, aviation or maritime?

2.8.2 Due to the various issues identified within this review that have arisen within the production of the VTT and VTTR estimates, it would be unwise to adopt the published results from the current study.

2.8.3 In the meantime, a query is raised in paragraph **Error! Reference source not found.** relating to the realism of the current freight : non-freight ratio for LGV journeys which is published in Table A1.3.4 of the TAG Data Book – this ratio should be reconsidered by the Department.

2.9 Approach to adjusting the new values over time

2.9.1 The recommended approach to adjusting the new values over time. The new study recommends the use of specific indices for the carrier component, such as wage growth or CPI, while recommending real GDP growth to uprate the shipper

component over time. Are these recommendations the most suitable basis for implementation?

2.9.2 While the study's recommendations for specific temporal adjustment indices appear to be reasonable, it would be best to consider them again after further evidence has been assembled to confirm the successful performance of the methodology in attaining the separation between vehicle related and product related VTTs.

2.10 Tests or further research?

2.10.1 Tests or further research that could provide further assurance over the results of the study.

2.10.2 It is not straightforward to see how the data available within this study could be successfully reanalysed to offset the various issues that have been identified above. The high level of uncertainty about the clarity of understanding by participants of the context of the experiments, coupled with sample size being below the quota level for 10 of the 12 segments (even before the removal of Shippers subset (B)) are major challenges for which effective solutions are not obvious. There may perhaps be scope to reuse this dataset with some creativity for OGV carriers, maybe making greater use of the substantial information collected in Part B of the questionnaire relating to the firms and their journeys, as well as through exploring different model forms to salvage some more firmly based results. Some suggested further analyses are outlined below in Section 7.4. However, is not easy for this reviewer to determine the likelihood of success from reanalysing the existing data from the experiments.

2.10.3 A number of recommendations are documented in detail in Section 8 below. These proposed adjustments to the current study approach should increase the likelihood of obtaining satisfactory estimates of VTT and VTTR in a future study. A larger initial sample size that is based on an improved segmentation into a more homogeneous set of vehicle types would improve the precision, cost-effectiveness and reduce the risks within a future study. Some improvements to the modelling approach would then become feasible through having a larger sample available for the analysis of each segment.

2.10.4 Three adjustments are proposed in Section 8.3 to the approach to the SP experiments and to the associated model estimation. These would enable the overall coherence of the approach to be tested more rigorously and would provide improvements in the precision of the estimates of VTT and VTTR.

2.10.5 The feasibility of carrying out revealed preference based studies to complement SP studies should also be considered. Examples are provided in Section 8.4 of the types of contexts in which revealed preference studies could be executed in the UK.

3 Why representativity is important for freight transport analysis

3.1 Introduction

3.1.1 The reason why the representativeness of the sample is of high importance for this study lies in the heterogeneous nature of freight requirements and responses. Experienced transport modellers will attest that behavioural/economic modelling of freight demand responses is inherently more challenging than the corresponding modelling of passenger demand. This is because of the high level of heterogeneity in requirements for freight transport between: different types of goods; when travelling within the different logistic stages of the supply chain; and within the different intermodal legs of a logistic stage. The Final Report confirms in section 3.1.2 that Technical Note 1.2 (TN 1.2) of the study “discussed the implications of heterogeneity of the goods”.

3.1.2 The nature and scale of this heterogeneity in road freight traffic is now illustrated, together with its implications for the measurement of VTT and VTTR values for specific types of HGVs operating in specific contexts of usage.

3.2 VTT and VTTR for HGV journeys to port

3.2.1 For example, the requirement to meet a specific time deadline at a port will be very high for an HGV that is travelling to board (i.e. **accompanied HGV**) a ferry service that has only a daily frequency, because missing this sailing would involve a day’s delay. Here the cost of travel time unreliability would be high because it could waste a day of driver’s wages, plus the daily capital cost of the tractor plus trailer, plus any cargo delay cost.

3.2.2 In contrast, the same commodity being delivered on the same HGV to the same port but instead for an **unaccompanied HGV** (i.e. trailer only) sea crossing service may have a much lower value of travel reliability. Unaccompanied trailers will be delivered to the port relatively continuously over time, rather than having their movements being strongly focussed to just before the actual ferry sailing time. Even if the preferred sailing is missed due to some road delays, it would only be the delay cost components of the capital cost for the trailer time, plus the cargo delay cost that would be incurred. Moreover, the value of goods and the perishability rate for those goods on unaccompanied HGV mode is typically significantly lower than that on accompanied HGV mode, due to the different commodity mixes that these modes transport to ports.

3.2.3 The delivery to ports of bulk goods, such as grains or construction materials, is likely to be even less time sensitive because these bulk goods are normally unloaded from the HGV to a silo or to some other storage facility from which the incoming ships are later loaded. HGV deliveries of containers to ports similarly tend to have considerable container dwell times at the port before being loaded onto the container vessel.

3.2.4 Due to the impact of Brexit, the balance in movements to UK ports has changed in recent years so that since 2021 unaccompanied HGV movements now exceed accompanied HGV movements³. When adding in the further many HGV deliveries to ports of containers and of bulk goods, it becomes clear that accompanied HGVs comprise only a minority of all HGV arrivals at UK ports. The implications of this should be recognised accordingly when estimating and expanding the VTT and VTTR of HGV journeys to ports.

3.2.5 In summary, the VTT and especially the VTTR for an HGV journey to a port is likely to be much greater on average for the minority accompanied HGV mode, than for either the majority unaccompanied HGV mode or for deliveries of containers or of bulk goods to ports. This differentiation would need to be recognised within the stated preference experiments and in the subsequent sample expansion procedure.

3.3 VTT and VTTR differentiation by commodity type, etc.

3.3.1 The next example focusses on response heterogeneity related to the type of goods carried. An articulated HGV carrying a full load of cigarettes or of scotch may have a capital value of its goods of up to a million pounds. Accordingly, the capital cost of delays and the costs of any insecurity in delivery will be large for such cargoes.

3.3.2 At the other end of the spectrum, many bulk inputs to and outputs from the construction sector may have very low capital value (e.g. stones, sand or gravel). In the particular case of the waste sector, the large tonnage transported by road may have a zero/negative capital value, where this waste is not suited to recycling.

3.3.3 This set of examples indicates just some of the ways in which the magnitudes of the VTT and VTTR are highly heterogeneous across the contexts of commodity type and of stage in an intermodal chain or in a supply chain. Accordingly, the average VTT/R for a specific type of vehicle will be strongly influenced by the mix of freight movement types that it carries.

3.3.4 The CSRGT survey of HGV usage in Great Britain indicates that the specific individual type and size classes of HGVs:

- carry quite different mixes of commodity types;
- cater for different logistic stages within supply chains and within intermodal chains;
- have very different likelihoods of use for journeys to ports or airports;
- exhibit different patterns of use across different types of roads;
- exhibit different patterns of use across times of day and days of the week.

3.3.5 This differentiation in usage between HGV classes, in turn strongly influences the values that would be expected for the VTT and VTTR associated with

³ See DfT Statistics Table PORT0103.

each such HGV class. It also highlights the importance of ensuring that appropriate observed vehicle mixes across HGV classes are applied across contexts within the procedure that expands the observed response values of the sample, up to provide representative averages across the population of HGV activity in Great Britain as a whole.

3.3.6 In the modelling of passenger transport, it is standard to apply different VTT values to different travel purposes. Moreover, a further differentiation between income groups is helpful when this distinction is feasible within a passenger model. In freight modelling the necessity for differentiation of VTT and VTTR values between homogeneous contexts is even stronger, for the reasons that have been illustrated by the evidence assembled in this note.

3.3.7 We next examine the extent to which these issues of road freight heterogeneity across HGV classes have been successfully addressed within the procedures adopted within this freight value of time study.

4 Selection of the sample

4.1 Sample design and potential for bias

4.1.1 This Section analyses the extent to which the sample of firms that were selected for the stated preference experiments for the freight value of time study are representative of road transport activity across Great Britain. The aim is to ascertain whether or not there are any substantial biases within the sample of freight industry respondents that was used. Such biases would complicate the procedures subsequently required in order to expand the sample-based observed estimates of the freight Value of Travel Time (VTT) and Value of Travel Time Reliability (VTTR) to make them representative of road freight activity patterns across Great Britain as a whole.

4.1.2 This analysis of potential bias is carried out through a number of steps. Firstly, it examines the manner through which the sample set of shippers (cargo traders) and the sample set of carriers (transport providers - hauliers) has been selected. The aim is to confirm that the combined sample set of participants adequately covers in an unbiased fashion the spectrum of actors across the road freight system of Great Britain.

4.1.3 The recommendations presented in FR Section 3.1.2 that were derived from TN1.2 of the study include:

“Further observations that freight transport is heterogeneous and can differ by value density, perishability of goods, shipment size, trip distance and time, and the impact of delay. Segmentation proposed in TN 1.1 and testing dependence on commodity type responds to heterogeneity. The study should also aim for a mix [in] responses from firms of different sizes.”

4.1.4 This text correctly identifies many of the major sources of heterogeneity in freight demand and in the likely resulting systematic diversity of responsiveness to transport cost or time savings. However, it does not clarify whether any of these specific known sources of heterogeneity were actually considered explicitly within the design and selection of the sample of respondents for the Stated Preference (SP) experiments. The subsequent discussion in FR Section 5.1 appears to imply that none of these sources impacted in practice on the design of the sample selected.

4.1.5 The sample profile that was aimed for within the chosen target of 600 interviews is discussed in FR Section 4.2 onwards. It aimed to achieve the segmentation specified in FR Table 5, which comprises combinations of the following three segmentation dimensions: vehicle type (OGV1, OGV2 or LGV) [3]; trips to ports, or other trips [2]; shippers or carriers [2]. This generated 12 [=3*2*2] combination segments, for each of which minimum quota of 50 SP interviews is required to successfully estimate a suitable logit model. This generates an overall minimum sample size requirement for at least 600 interviewees to successfully complete the whole of the online survey.

4.1.6 During the creation of the sample, they “targeted the ports and maritime sector as a mid-survey review of respondents showed that this sector was under-represented. This technical input provided a significant increase in responses from this sector.” (FR p.13). However, this port targeting is likely to have added bias to the sample as a whole, particularly because only a very small proportion of all carriers, other than those heavily using large articulated HGVs, are likely to generate regular movements to ports or airports.

4.1.7 The size of the survey response rate provides a useful indicator of the potential scope for bias to occur within the cross-section finally captured within a sample. As documented in FR Table 6 for the main SP experiments, after putting aside the 7,357 non-contacts there were a further 2,769 who refused to participate, whereas 6,132 agreed to take part and so were sent unique links to complete the survey. However, only 602 actually then completed the survey. This implies a response rate below 10%. This is not ideal because the set of characteristics and the cost and time sensitivity rates of the 90+% who did not respond, *might* be quite different to those of the respondents. Significant effort would be required in order to examine and then mitigate any such *potential for* bias, through the design of a suitable procedure that expands from this sample up to match the structure of the GB population of shippers and carriers. It does not appear from FR that this aspect has been addressed in practice.

4.1.8 The sampling profile that was actually achieved is reviewed in FR Section 4.5, which provides an informative overview of the degree of match of: the cross-section captured by the sample of 602 respondents; to that expected for Great Britain as a whole. While these matches are broadly encouraging on some metrics, FR Table 8 indicates that there is a major under-representation of small firms, because only 1% of the sample firms had less than 10 employees. The IDBR data for 2023 published by ONS (see Table 1 below) indicates that in reality in the UK, these small firms comprise 93% of the total road freight transport firms, which indicates

the scale of the mismatch between the sample composition and the UK road haulage industry as a whole. 91% of the sample was for firms of 50 plus employees, whereas just 1% of UK haulage firms are in this size band, though due to having many vehicles these large haulage firms will carry substantially more than 1% of GB tonne kilometres. Clearly the distribution of firms by size within the sample cannot be deemed to be representative of the UK population of hauliers.

Table 1 Number of VAT and/or PAYE based enterprises by employment size bands for freight transport by road, UK 2023

Employment Size Band							Total
0-4	5-9	10-19	20-49	50-99	100-249	250+	
45,065	5,230	2,515	1,165	365	180	75	54,595
83%	10%	5%	2%	1%	0%	0%	100%

Source: ONS Table 4 (2023) using IDBR data for SIC class 4941,

<https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/data/sets/ukbusinessactivitysizeandlocation>

4.1.9 The discussion on FR p.16 of the reasons behind the under-representation of smaller businesses within the sample of carriers seem plausible. It illustrates one potential substantial source of bias within the sample, together with identifying some of the side-effects that this causes regarding the consistency of coverage across the spectrum of product types and regarding the resulting bias towards higher value product types within the sample.

4.1.10 In the informative comparison within FR Table 10 that presents the spread across product types from the sample, versus that from the CSRG, it would have been helpful to also present the CSRG percentage splits in units of journeys, rather than just in units of tonnes and tonne kilometres. Product categories such as: Mail and parcels; or Retail (non-food), tend to be relatively light in weight, so part of the apparent tonnage over-estimation within the sample that was presented would be offset by this change to adopt the more directly comparable vehicle units as the basis for cross-checking. Because this study of VTTs is primarily designed to support road traffic analysis and assessment for both TAG / DfT and National Highways, rather than to support economic or industrial analysis for other sections of government, it is the number of vehicle units and of vehicle kilometres that is of central relevance, rather than the number of tonnes lifted or of tonne kilometres moved.

4.2 Sample quota achievement

4.2.1 Table 2 summarises how the 572 SP respondents that remained after 30 were removed⁴ during the data cleaning process, were distributed across the different combination subsamples (i.e. carriers vs. shippers, shipments (not) to port, and vehicle type). It lists the desired sample quota totals in parentheses for the row and column aggregates, noting that each individual combination had a desired quota of 50.

Table 2 Number of responses, (Quota) for different vehicle types, destinations by shippers and carriers

		OGV2	OGV1	LGV	Total
Shippers	To port	34	27	33	94 (150)
	Not to port	18	63	30	111 (150)
Carriers	To port	52	50	24	126 (150)
	Not to port	45	145	51	241 (150)
Total		149 (200)	285 (200)	138 (200)	572 (600)

Source: FR Table 12

4.2.2 This sample was used to carry out the SP experiments. It comprised a total of 367 carriers and 205 shippers, rather than being evenly split with 300 for each. There is a further imbalance of 208 OGV1 vehicles **not** going to port, versus 77 going to port, instead of the desired quota of 150 for each. Overall, 4 of the 6 combinations for carriers achieved or exceeded their quota of 50, whereas 5 of the 6 shipper combinations are at least 32% below their quota totals.

4.2.3 Section 5.1 below discusses the application after the SP experiments were completed of the exclusion criteria that further removed from the sample those responses that were implausible. This step in turn led to further reductions below quota for many segments and generated a sample size of just 472 within the model estimation stage, rather than the envisaged 600 sample size. As discussed further in the Review sections below, this failure to achieve quota totals impacted on the ability of the study to estimate VTT values at the desired level of segmentation and it may also have prevented the determination of appropriate external explanatory variables for use in rebasing the sample to mitigate the impacts of sample bias.

⁴ Most of those 30 removed at this stage were those shippers with their transport only on own account. For reasons outlined below in Section 7.1, these needed to be removed from the modelling analysis step to enable the resulting VTT to refer only to product related and not transport related costs.

5 Design of experiments

5.1 Exclusion criteria

5.1.1 FR Section 5.1⁵ documents the five exclusion criteria checks carried out on the results provided by the SP experiments. These checks provide useful guidance to confirm the logic of the results and to confirm that the respondents were confident about their ability to make sensible and informed choices between the alternatives that were presented to them in their SP experiments. The analysis of these distinct checks is an important and useful way of reducing noise and/or errors in the results and in guiding on areas of potential concern within the experiments. An informative discussion of the findings has been presented.

5.1.2 As well as checking the impact on the sample size of each of the exclusion criteria, simple choice models were estimated to determine the impact of each exclusion criterion on the estimated VTT and VTTR values (as documented in FR Appendix A.3). The balancing of the reductions in sample size against the generation of more plausible VTT and VTTR values, led to a decision to only apply two of the five exclusion criteria, as follows.

- **(C1) Removal of non-rational respondents**, was applied fully.
- **(C5) Removal of respondents accepting a very high boundary value of travel time (BVTT)** was applied selectively, focused on where it appeared to improve the plausibility of the resulting VTT and VTTR estimates.
- Whereas **(C2), (C3) and (C4)** were not applied. These three criteria relate to the respondents' reported understanding of the clarity or consistency of the context or questions presented to them within the SP experiment.

5.1.3 The criterion: **(C4) Removal of respondents perceiving the description of the journeys to be unclear**, had the largest failure rate of the five criteria. This exclusion criterion asks respondents with regard to their SP experiment whether the descriptions of the journeys under analysis were clear. All respondents answering 'no' would be excluded under this exclusion criterion which would potentially have excluded 195 of the original 572 sample within the experiment stage.

5.1.4 The later discussion at the end of FR Section 5.3 together with the FR Table 27⁶ provide further context. Firstly, the three questions D12 to D14 that determined the exclusion criteria: (C2 – D12), (C3 – D14) and (C4 -D13) were asked only after both the SP1 - VTT and the SP2 - VTTR experiments had been completed. Consequently, it is not discernible whether any recorded lack of understanding should relate primarily to the more complex SP2 reliability experiments or whether it equally relates to the SP1 VTT experiments. Secondly, these questions D12 to

⁵ Further discussion and tabulation of these exclusion criteria results is provided subsequently at the end of FR Section 5.3 and in its Table 27. It would be clearer to the reader if these two distinct sections had been unified into a single integrated presentation.

⁶ Within this FR Table 27, the Table reference numbers have mistakenly switched "D14" and "D13".

D14 were optional. Around 64% responded to D12 and to D14 but just 46% responded to question D13 for exclusion criterion (C4). More specifically for (C4), only 57 (12%) responded that the journey descriptions were clear, whereas 160 (34%) said they were not clear, while 255 (54%) did not respond at all. It is not an encouraging indicator to find that only this small minority responded that the descriptions were clear!

5.1.5 The two exclusion criteria (C1) and (C5) that were actually applied caused further reductions down to 472 in the effective sample size available for use within the model analysis stage, as illustrated for the sample combinations plus quota below in Table 3. Only 2 of the 12 segments, namely the OGV1 shippers and OGV1 carriers not to port, have achieved their quota of including at least 50 respondents. The largest sample reduction was from 138 to 85 for LGVs, which was mainly due to exclusion criterion (C5), for the reasons discussed later at the end of Section 5.2.

Table 3 Sample sizes used for analysis, (Quota) for different vehicle types, destinations by shippers and carriers.

		OGV2	OGV1	LGV	Total
Shippers	To port	31	26	22	79 (150)
	Not to port	18	55	18	91 (150)
Carriers	To port	47	49	19	115 (150)
	Not to port	39	122	26	187 (150)
Total		135 (200)	252 (200)	85 (200)	472 (600)

Source: FR Table 14

5.1.6 Conclusions on two important issues can be drawn from this analysis of the sample of respondents eventually used within the model analysis stage. Firstly, the major sample size imbalances and the resulting sample sizes falling below quota totals, will each impact on the effectiveness of the steps of experimental model analysis and of the expansion of the sample results to represent Great Britain as a whole. Secondly, the non-application of the three exclusion criteria (C2) to (C4) indicates that at least a quarter, and possibly over a half, of those eventually included within the SP experimental results had indicated reservations regarding either the realism and/or their understanding of the context or questions that were used within their experiments. Unfortunately, these two issues interact with each other. Any increased rate of exclusion of those respondents who report reservations on their understanding, would in turn tend to magnify the reductions below quota in the available sample of respondents.

5.1.7 These two issues should not be viewed as an indicator of sub-standard performance by those designing and carrying out the SP experiments. They are more a consequence of the complexities that arise when carrying out what is a methodologically challenging study within a limited budget available for sampling. If it had been feasible to start instead with a sample, say, of 1,800 respondents, then it should have been feasible to circumvent these two issues, without major changes from the broad approach that was adopted.

5.1.8 In general, FR section 5.2 provides very useful tables and background information about the respondents' typical transport shipment and journeys.

Nonetheless, the presentation of some further details would have been informative. The useful FR Tables 15 to 17 would benefit from being segmented between OGV1 and OGV2. Likewise, LGV tables that correspond to the OGV tables 16 and 17 would be useful to include. The statement in FR bottom p.20, that “For LGV type vehicles only a handful of shipments went to respectively rail- and airports” does not make explicit whether many or whether zero LGV shipments went to seaports.

5.1.9 It would also be helpful to see for the set of tables 15 to 22 a comparison of: the number of respondents in the SP sample actually used for model analysis; versus the numbers remaining after all five exclusion criteria have been applied to this SP sample. These extended tabulations would help readers to understand the scale and types of bias in the estimates that could arise due to particular segments of respondents reporting uncertainty regarding the clarity of the questions and/or context within the experiments.

5.2 Transport cost relativities and BVT levels in SP1

5.2.1 The transport cost values arising within this study generated concerns relating to two interdependent issues

- a) unexpected relativities between vehicle types for the typical costs per journey hour that were reported by respondents
- b) major differences between individual segments in the ratio of:
 - the Boundary Values of Travel Time (BVT – which in effect is an implicit cost per hour) option sets introduced into their SP experiment;
 - compared to the typical cost per hour as reported by the respondents from that segment.

These two issues are further discussed in turn below to explain them more fully.

5.2.2 The first issue has been identified as follows.

“Reported transport costs for OGV1 were, however, consistently higher than for LGV, and OGV2 transport costs were relatively low. (FR p.22)

As mentioned in Section 4, the obtained set of goods transported was not representative for the population.” (FR p.20).

5.2.3 It is unexpected that the comparison of respondent’s reported transport costs per hour (FR Table 20) indicates cost relativities of OGV1>LGV>OGV2 for each of carriers and of shippers. This sequence differs from the expected driver and capital cost relativities for these vehicle types, namely OGV2>OGV1>LGV and accordingly differs from the operating cost values that are specified for these vehicle types in TAG. It may well imply a major bias relative to the overall UK population of firms, in the characteristics of those sampled for some or all of the vehicle type segments. The need to find firms that have regular shipments to ports/airports for OGV1s and LGVs may have contributed to such bias. Any such sample bias would impact on the validity of the results from the subsequent analysis unless some convincing method could be introduced to offset that bias, as discussed later in Section 7.4.

5.2.4 There is a further cost anomaly, that the average shipment cost for a single journey leg that was observed within the pilot stage was £253 (FR p.9), whereas the

corresponding mean costs per journey subsequently observed within the main sample (FR Table 20) varied from £69 for LGV carriers to £179 for OGV2 carriers. This appears to suggest that the sample in the pilot stage was substantially different from that in the main SP experiment stage.

5.2.5 Turning now to the second of the issues listed above, FR p. 23 explains that the cost and time values used as inputs to the SP1 experiments that were carried out with the main sample, were based on the self-reported characteristics by the respondents from this main sample, rather than on those from the pilot sample. However, the observed pattern of average transport costs per hour discussed above in paragraph 5.2.3, varies in a completely different pattern, across shipper/carrier and across vehicle types, to the pattern of average Boundary Values of Travel Time (BVTT) presented to respondents in the SP1 experiments, as summarised in FR Table 25. This Table is reproduced below here as Table 4 but now in a revised format in order to present various aspects in a more informative arrangement. The implied set of BVTT values that were indirectly presented to the respondents of a segment within their 8 SP1 experiments are now presented in Table 4 as a proportion of the mean cost per hour observed for that specified segment, rather than being presented in absolute values, as originally in FR Table 25.

Table 4 For each segment: Observed mean journey cost per hour; for each bid rank - the average BVTT presented to respondents now illustrated as a proportion of this mean cost

	OGV1		OGV2		LGV	
	Shippers	Carriers	Shippers	Carriers	Shippers	Carriers
Mean cost/hour ⁺	£74.75	£74.84	£55.01	£66.49	£64.54	£66.56
BVTT presented for specified rank, as a percent of observed mean cost per hour						
Lowest bid	3.8%	24.4%	12.3%	54.4%	1.2%	4.7%
2nd lowest bid	6.4%	40.5%	20.3%	89.5%	2.2%	8.4%
3rd lowest bid	9.4%	60.1%	29.6%	133.9%	3.3%	12.6%
4th lowest bid	15.4%	93.4%	49.6%	220.3%	5.5%	20.2%
4th highest bid	26.9%	149.1%	89.1%	335.7%	8.8%	29.2%
3rd highest bid	46.3%	243.8%	145.0%	556.5%	13.4%	45.8%
2nd highest bid	75.3%	410.2%	231.1%	894.6%	22.2%	79.0%
Highest BVTT	166.8%	932.1%	489.0%	1801.6%	44.4%	150.9%

Source: FR Tables 25 and 20⁺

5.2.6 The colouring based on relative size in Table 4 highlights the systematic major differences between the segments in the patterns of their relative BVTT value option sets that were introduced into the experiments. This BVTT pattern imposed by the experimental design is systematically higher or lower than the corresponding mean cost per hour of that segment, following a pattern that depends on the segmentation by shipper/ carrier and by vehicle type. For example,

- the **highest** BVTT cost per hour option presented to LGV shippers is set to **be just 44%** of the observed mean cost per hour for this segment,
- the **lowest** BVTT cost per hour option presented to OGV2 carriers is set to **be 54%** of the observed mean cost per hour for this segment,

- whereas the **highest** BVTT cost per hour option presented to these OGV2 carriers is set to be **18 times larger** than the observed mean cost per hour for this segment.

5.2.7 The source of this large differentiation between segments appears to relate to the segment specific values of the scale factor *BVTTfac* that determine the BVTT-ranges in units of £/ton/hour, which are tabulated in FR Table 45, as reproduced here.

FR Table 45: BVTT-range for the different segments				
BVTT-range (£/ton/hour)				
Goods value	Low-value	Medium-value	High-value	Perishable
Carrier	1.0 – 74.0 (<i>BVTTfac</i> = 1.0)	1.0 – 74.0 (<i>BVTTfac</i> = 1.0)	1.0 – 74.0 (<i>BVTTfac</i> = 1.0)	1.0 – 74.0 (<i>BVTTfac</i> = 1.0)
Shipper (own-account)	1.1 – 81.4 (<i>BVTTfac</i> = 1.1)	1.15 – 85.1 (<i>BVTTfac</i> = 1.15)	1.25 – 92.5 (<i>BVTTfac</i> = 1.25)	1.4 – 103.6 (<i>BVTTfac</i> = 1.4)
Shipper (hire-and-reward, Light Goods Vehicle)	0.2 – 14.8 (<i>BVTTfac</i> = 0.2)	0.3 – 22.2 (<i>BVTTfac</i> = 0.3)	0.4 – 29.6 (<i>BVTTfac</i> = 0.4)	0.5 – 37.0 (<i>BVTTfac</i> = 0.5)
Shipper (hire-and-reward, Other Goods Vehicle)	0.1 – 7.4 (<i>BVTTfac</i> = 0.1)	0.15 – 11.1 (<i>BVTTfac</i> = 0.15)	0.25 – 18.5 (<i>BVTTfac</i> = 0.25)	0.4 – 29.6 (<i>BVTTfac</i> = 0.4)

Source: FR, p.61

5.2.8 These BVTT ranges were based on the most recent Norwegian Freight VTT study and from European research literature (FR p.61). Because carrier respondents were explicitly requested within the survey to avoid considering the impacts on the products they are carrying,

“The BVTT-range for carriers is not dependent on goods value, as for carriers the goods component is not part of the scope of the survey.” (FR p.61)

it is difficult to understand why a BVTT range that is in units of £/ton/hour has been applied to them (FR Table 45). The large differences in relative BVTT values between vehicle types indicated for carriers in Table 4, would appear to be a direct result of this use of average load per vehicle to scale the range. For carriers it would seem to be much more logical to instead adopt a BVTT range in units of £/vehicle/hour that is segmented by vehicle type so as to take account of standard differences in vehicle operating costs. This modification would circumvent the extreme differences in relative BVTT value ranges for carriers discussed in this Section.

5.2.9 For shippers, the setup is more complicated. The BVTT range is scaled (FR Table 45) to be between 3 and 10 times larger for those shippers who only use own-account transport (Shippers C), than for those shippers who only sub-contract to hire-and-reward hauliers (Shippers A). This differential scaling is aimed to represent the time costs of transport plus those of products for the Shippers C but only the time costs of products for Shippers A. However, in practice within the experiments the own-account Shippers C were excluded from the sample that was modelled. The largest part of the remaining shipper sample was the intermediate subset - Shippers B - who use some, but not solely, hire-and-reward hauliers. It

does not appear to be clarified in the Final Report whether Shippers B have been allocated the BVTT range associated with Shippers A or with Shippers C. Based on the response guidance they received that advised them to consider both the time costs of transport plus those of products, when selecting between journey options, Shippers B should logically have been allocated the BVTT range for own-account Shippers C. Because the results for shippers above in [Table 4](#) represent Shippers A plus B combined, it is not straightforward to determine whether Shippers B were correctly allocated this BVTT range for own-account Shippers C. However, the comparison of the number patterns for LGVs between carriers and shippers appears to suggest that Shippers B had been allocated the BVTT range appropriate only for Shippers A, rather than adopting the correct Shipper C range.

5.2.10 The complicated discussion within the previous three paragraphs explains why there are such major systematic differences between segments in the relative BVTT ranges that they face and why much of this differentiation appears to have been erroneously generated. The discussion below focuses on the problems that these major systematic range differences have placed on the subsequent attainment of well estimated models of VTT and VTTR.

5.2.11 This extreme pattern of differences between segments suggests that many of the respondents in certain segments will have been presented with many BVTT option levels that differed substantially from their own expectations on mean costs. Perhaps this anomaly goes some way towards explaining the high proportions of respondents that struggled with the clarity or consistency of the context or questions in the SP experiment, as evidenced by the numbers failing the exclusion criteria (C2) to (C4) that have been discussed above in paragraph 5.1.5?

5.2.12 Relating to FR Table 23 that presented the “Number of times the fast and expensive alternative was chosen in SP1 by vehicle type”, FR states:

“Shippers were in general more drawn towards the fast but expensive alternatives compared to their carrier counterparts using the same vehicle type and this tendency increased as the size of the vehicle type decreased. Part of this may be driven by experimental design differences, since presented transport costs and times were varied based on the specifics of shippers/ carriers and the used vehicle type” (FR p.24)

5.2.13 [Table 5](#) below reformulates the FR Table 23 to use cumulative proportions and then uses the colouring by relative size format to illustrate that this table now has a very similar relative size pattern across segments to that found above within [Table 4](#). This again suggests that the apparent differences between shippers and carrier choices discussed in the FR were primarily driven by the extreme differences between segments that were explicitly introduced into the BVTT option sets within the SP1 experimental design. Specifically, where the presented BVTT costs are high relative to the respondents’ observed average cost (e.g. OGV1 or OGV2 carriers), most respondents avoid the fast and expensive options. In contrast, where the presented BVTT costs are low relative to the observed average (e.g. LGVs or OGV1 shippers), most respondents choose the fast and expensive options.

Table 5 Cumulative percentage of number of times fast and expensive is chosen

Number of times	OGV1		OGV2		LGV	
	Shippers	Carriers	Shippers	Carriers	Shippers	Carriers
0	0%	1%	2%	14%	0%	0%
1	0%	7%	2%	37%	0%	2%
2	1%	22%	2%	75%	0%	2%
3	3%	39%	8%	89%	5%	4%
4	13%	59%	28%	95%	8%	8%
5	24%	81%	40%	97%	8%	15%
6	45%	92%	64%	98%	23%	28%
7	70%	98%	88%	100%	56%	70%
8	100%	100%	100%	100%	100%	100%

Source: FR Table 23 reformulated to cumulative proportions.

5.2.14 FR states that

“The experimental design used the self-reported transport costs and travel times for the reference transport as the reference point. Positive and negative variations in transport costs and travel times were presented around these reference values. The ranges in transport costs and travel times were based on existing knowledge of VTT values obtained in similar SP studies in the international literature ... The empirical literature suggested that VTTs for shippers were lower than carriers, and hence shippers are presented with lower BVTT values. Similarly, the VTT was expected to increase with the size of the vehicle (due increased costs of drivers and depreciation etc.) and the design therefore presented higher BVTT values for larger vehicles.” (FR p.25).

5.2.15 The results above in Table 4 presenting the BVTT relative to the observed mean cost per hour (the reference point), indicate that the average pattern of BVTT variations presented was systematically skewed to either be largely positive or largely negative for individual segments. For OGV2 carriers, for example, 6 of their 8 BVTT options exceed their reference point, whereas for LGV shippers all of their 8 BVTT options are set below their reference point. The various oddities discussed above from paragraph 5.2.7 onwards, appear to indicate that the range of BVTTs actually used for some of the segments, may have been adopted in error.

5.2.16 With respect to the results presented in FR Table 24 for the “Share of respondents accepting the BVTT by rank“, FR states that

“The designs were tested in the pilot, during which the rate of accepting the highest BVTT values was not as extreme as in the present sample. The relative differences between shippers and carriers in presented BVTT values (a factor 5-6) may explain the discrepancies across shippers and carriers. *However, the observed response patterns appear a specific feature of the sample, for which we do not have a clear explanation.*” (FR p.25, italics added)

Table 6 For each segment, the % acceptance of BVTT when set: to 70% of its mean cost; or to the highest presented BVTT

	OGV1		OGV2		LGV	
	Shippers	Carriers	Shippers	Carriers	Shippers	Carriers
Mean cost/hour ⁺	£74.75	£74.84	£55.01	£66.49	£64.54	£66.56
% who accept a BVTT that is set to 70% of the observed mean cost for the segment	67.9%	62.3%	78.3%	58.5%	<68%	78.5%
% who accept the highest BVTT presented to that segment	52%	15%	22%	3%	68%	49%

Source: FR Tables 20, 24 and 25.

5.2.17 The results derived from FR Table 24 that are presented in the middle row of Table 6 were calculated by adopting an arbitrary benchmark BVTT value that is set for each segment to 70% of that segment's observed mean cost per hour, then interpolating between the values in the FR Table 25, based on the percentages in FR Table 24, to derive the proportion that would accept this BVTT value. It is instructive that the resulting acceptance rate for the BVTT when it is set to this benchmark 70% value, is clustered within the interval of 58% to 78% of respondents across the 6 segments. This would appear to suggest that the differences between the segments in their stated preferences for BVTT are not so great, despite the radical differences in the option sets presented to them within the SP1. It provides some support for the view that an experiment that provided appropriate BVTT option sets for all segments would be likely to perform more plausibly than did the current set of erratic ranges.

5.2.18 In contradiction to the viewpoint from the FR p.25 that was presented in italics above, this combined evidence strongly suggests that there actually is a "clear explanation" for the observed response patterns, namely that it is a direct result of the peculiar design of the experiment, rather than being "a specific feature of the sample".

5.2.19 The final row of Table 6 presents the percentage who accept the highest BVTT that was presented to that segment. This percentage has a pattern of major differences (from 3% to 68%) across segments that is similar in its relative magnitudes to the pattern by segment illustrated in the final row of Table 4 above, for the (inverse of the) highest presented BVTT as a proportion of its observed mean cost per hour. This percentage who accept the highest BVTT are problematic within the estimation process.

"This response pattern drove up the VTT since the models were unable to put an upper limit on the VTT for the respondents always selecting the fast and expensive option." (FR p.33)

5.2.20 To mitigate this problem, which was pronounced for all shippers segments as well as for LGV carriers, it was decided to implement the preferred mixed logit models estimated using a lognormal density only for carriers but not for shippers, in order to circumvent obtaining very high VTT estimates for shippers.

5.2.21 For similar reasons - “the SP choice data revealed that VTT values are coming out rather high” (FR p19) - the exclusion criterion (C5)-**Removal of respondents accepting a very high boundary value of travel time (BVTT) in SP1** was introduced. However, it was introduced as an upper bound *per tonne* per hour, for both shippers and carriers, despite the experiment’s requirement for carriers to ignore costs related to products which implies that a carrier cut-off *per vehicle* per hour would have been more appropriate. FR Table 13 indicates that this exclusion criterion (C5) led to at least 43 out of the original sample of 138 LGV respondents being removed due to exhibiting very high BVTT values. This in turn caused the LGV sample size for each individual segment to fall below half its required quota. This of course had knock-on effects in reducing the ability to estimate suitable LGV choice models. This set of problems appears to have been an unfortunate side-effect generated from the original selection of inappropriate BVTT ranges for the LGV shipper and carrier segments.

5.2.22 In summary, the pattern of BVTT ranges that were presented to respondents within the SP experiments were radically different in their relative cost magnitudes across the set of different segments. In many cases the scale of these differences appears to have arisen due to obscure design mistakes / misinterpretations, rather than for productive reasons. These have resulted in modelling problems being generated for the shippers segments and for LGV carriers, as well as to a loss of sample size due to the exclusion criterion (C5) needing to be applied. They may also have tended to bias upwards the estimated VTTs.

5.3 SP2 - reliability experiments design

5.3.1 The analysis within the SP2 experiments of the influence of reliability, necessarily introduces a further element of complexity to the design of the experiments because it requires the introduction of a third dimension: the variability of travel time, in addition to the original average time and cost dimensions for a journey.

5.3.2 The set of issues discussed in the previous Section related to the SP1 experiments that determine the VTT, will also impact on the estimation through SP2 experiments that determine VTTR. This is because the choice results data from these two experiments is pooled as input to the model which simultaneously estimates both VTT and VTTR.

5.3.3 FR Table 26 appears to have an error for the OGV2 tabulations as its columns each sum to 60%, rather than to 100%. This error, coupled with the lack of inclusion in FR of further explanatory tabulations for the SP2 experiment, which would be analogous to the FR Tables 23 to 25 for the SP1 experiment, makes it difficult to assess whether or not similar design issues to those raised for SP1 within the previous section, have also arisen explicitly for SP2.

6 Analysis of experiments

6.1 ISSUES FOR FURTHER CONSIDERATION

6.1.1 The modelling approach that was used to analyse the results from the SP experiments to derive VTT and VTTR magnitudes, is presented in FR Section 5.4. Three main theoretical variants on the choice model form were explored. The estimates that were produced from this analysis were discussed in FR Section 5.5 for the set of different model forms that were investigated.

6.1.2 Initially various choice model forms were investigated and based on the comparison of the resulting estimates of goodness of fit the most suitable model forms were adopted for specific segments. Then initial coefficients and standard errors were estimated for versions of these models that included a full set of explanatory variables. Next, the models were streamlined by systematically removing those independent variables that were not significant, until the finalised model form was achieved. This general approach is appropriate but perhaps because the available sample size for most segments was smaller than required, several oddities emerged in the results produced by these explorations. These are now discussed in turn.

6.1.3 Few but odd significant explanatory variables. When testing the OGV model, care was taken to initially include a wide range of potential external factors that might be expected to influence the VTT values, such as: type of goods, whether to port, and many others. However, for the OGV shippers' model (FR Table 31) just two such variables were initially significant at the 5% level but both had signs reversed from that expected. Neither the VTT nor VTTR parameters were significant at the 5% level in the OGV shippers initial full model. When the OGV shippers model was gradually streamlined through removing insignificant variables, this streamlining procedure led to those variables that had initially been significant subsequently losing their significance level, presumably in part due to correlation across some of the explanatory variables. In contrast, the retained VTT and VTTR parameters became significant at the 5% level or better, within this streamlined model form for OGV shippers (FR Table 32). The multiplier for trips to port was not statistically significant even at the 10% level but nevertheless it was retained within the streamlined shippers model due to its explicit role within the study. The OGV carriers' model with its larger sample size had three significant explanatory variables initially (FR Table 34) but within the OGV carriers model streamlining procedure these three were later excluded because their magnitudes were not deemed credible.

6.1.4 **VTTs for OGVs reversed.** The bottom of FR p.33 contrasts the VTT results generated from the Random Value (RV) model with those from the relative model. For the VTT from the relative model it found "For OGV2 trip, these values were 80-100% higher than their RV counterparts." Even though for carriers (but not for shippers) the relative model of OGVs provided a better fit than the RV model, "the recommended approach was to adopt the RV model specification with multiplicative

error terms for both shippers and carriers” (FR p.34). It would be interesting to investigate whether the subsequent findings for OGVs during the RV model estimations for both shippers and carriers that the VTT for OGV2 was unexpectedly⁷ smaller than that for OGV1 (FR p.36 and p.39), might have been reversed through switching to instead use the relative model formulation? An alternative approach was adopted to circumvent the anomalous estimated VTT relativities between OGV1 and OGV2, through forcing their estimated values to be identical, within the finally estimated streamlined models for each of shippers and carriers.

6.1.5 Small sample size for LGVs. The extract below explains why large numbers of LGV respondents were excluded from the analysis sample, namely 20 LGV shipper and 30 LGV carrier respondents were removed, thus losing more than a third of the initial LGV sample of 138 that was achieved.

“The analysis is primarily focused on the OGV sample. The reason for this is that during preliminary analyses the LGV sample was displaying very high VTT values”. (FR p.32)

“The LGV shippers sample was more problematic, and the assumptions in relation to exclusion criterion C5 were largely influential. Since LGV shippers were consistently accepting high BVTT bids in SP1, the corresponding VTT was much larger than for the OGV estimates. ... The displayed choice behaviour by the LGV shippers together with the small sample size led to the decision not to base the corresponding VTT estimates based on the analysis for this subsample. Instead, the recommendation was to apply the OGV shipper VTT and VTTR estimates to the LGV shippers cohort.” (FR p.41)

The discussion in Section 5.2 above has indicated that the major losses from the sample of LGVs, particularly for shippers, appears to be an indirect result of the low values for BVTT that were allocated within the option set in the experimental design for the LGVs.

6.1.6 Non-application of exclusion criteria. These models were estimated on a sample that included many respondents that had failed the exclusion criteria (C2) to (C4), having instead only excluded those failing criteria (C1) and partially (C5). The potential impacts on model results of applying these exclusion criteria are explored in FR Appendix A.3. However, the model form that was used within this early exploration is different in many ways to that subsequently used for the final calculation of the VTT and VTTR. As a result, the pattern and level of many of the estimated VTT values across segments presented for the finalised models in FR Table 40 are very different to those resulting from applying the exclusion criterion (C1) in the earlier exclusion criteria tests, as illustrated below in [Table 7](#).

⁷ OGV2 typically has higher driver and running costs than OGV1, so that time savings for OGV2 should not be less valuable than for OGV1.

Table 7 Estimated VTT by segment from early test of exclusion criterion (C1) and from finalised model

		LGV		OGV1		OGV2	
		Early	Final	Early	Final	Early	Final
Shipper	To port	£225	£95	£118	£95	£95	£95
	Not to port	£183	£72	£96	£72	£77	£72
Carrier	To port	£337	£64	£114	£81	£62	£81
	Not to port	£400	£64	£136	£81	£74	£81

Source: FR Tables 40, 63 and 64

6.1.7 This Table highlights that for the model used in the early exploratory tests, the magnitude of its estimated VTTs

- for LGVs, are between 2 to 6 times larger than those derived from the final model and so unexpectedly they are much larger than those for OGV1 or OGV2
- for OGV1s, are up to 68% larger than those derived from the final model and unexpectedly they are larger than for OGV2
- it is only for OGV2s that the VTTs are of a similar magnitude to those derived from the final model
- for the carriers segment only, for each vehicle type the VTTs are larger for the segment: not to port, than for the segment: to port. This is the reverse of expectations.

6.1.8 The magnitude and ubiquity of these differences in results between the early exploratory model and the final model does not provide confidence in the usefulness of these early exploratory estimates of the potential impacts on model results of (not) applying the individual exclusion criteria. It would have been informative instead to have carried out these sensitivity tests within model formulations that were similar in their mathematical form and in their range of independent variables to those used in the finalised models.

6.1.9 Response to unexpected intermediate results. The discussions above have outlined a number of instances in the course of the model analysis phase where unexpected results emerged for the estimated parameters. Such situations should act as a warning of the need to reexamine the fundamental procedures that gave rise to the data generating these errant results. For example, is the sample fully representative or has it biases that need to be mitigated? Have processing errors been introduced along the way that would need to be corrected?

6.1.10 Unfortunately in order to maximise the chances of achieving results that are broadly consistent with other past studies, the approach in this study often has tended instead:

- either to censor the data - e.g. by applying exclusion criterion C5 to remove part of the sample, instead of searching to locate the inappropriate design of the experiment that generated an erratic pattern of BVTs across segments, which in turn generated the large failure rate for C5;

- or to censor the modelling procedures - e.g. by removing significant explanatory variables with wrong signs, or by reformulating the model so as to force the OGV2 VTT to be no lower than that of OGV1, instead of ensuring that the initial sample of firms was unbiased and was sufficiently representative of the population to guarantee that the relativities for observed average journey costs per hour were plausible across vehicle types.

6.2 SUMMARY OF ISSUES

6.2.1 For the model analysis tasks, the relatively small sample sizes that fell below quota for many of the segments led to complications within the model estimation procedures. To resolve these issues, segments were combined so as to provide the VTT and VTTR results in FR Table 39, which distinguished just the following four aggregate segments, rather than the originally planned 12 detailed segments.

- OGV shippers (i.e. OGV1 and OGV2 combined), further segmented by to port / not to port, though using a parameter that was not statistically significant.
- OGV carriers
- LGV carriers.

In the absence of a convincing estimated model for LGV shippers, it is assumed that LGV shippers should be allocated the same VTT and VTTR as OGV shippers.

6.2.2 For both shippers and carriers the initial estimates of VTTs for OGV1 were larger than for OGV2, which was contrary to expectations. Accordingly, the estimated VTT and VTTR were then forced to adopt common values across the OGV1 and OGV2 segments.

6.2.3 It has not ultimately been feasible to identify the impact on VTT or VTTR of any external explanatory variables (e.g. commodity type or value, company size, etc.), though this lack of identified impacts may result mainly from the small samples available, rather than automatically indicating that in reality there is no substantial impact. Some initially significant explanatory variables were removed from the OGV carriers model due to assumptions that their magnitudes were not credible. This absence of external effects also precludes any post-processing of the results in order to remove potential biases due to lack of representativity in the original sample.

6.2.4 To avoid suffering even smaller samples, the final models were estimated using samples that included substantial numbers of respondents who ideally would have been dropped due to their failing the exclusion criteria (C2) to (C4). There is some limited evidence that this may have biased upwards the VTT values, while it certainly will have reduced the robustness of the evidence underlying the parameter estimates.

7 Usage in TAG

7.1 ORIGINAL APPROACH AND ITS ISSUES

7.1.1 FR Section 5.6 explains how the parameter values estimated from the models were transformed into VTT and VTTR parameters suitable for inclusion within TAG. In the absence within the models of credible estimates of any external influences on the VTT or VTTR values, there was no scope to mitigate the sampling bias impacts on these parameter values, which arose from the lack of representativity of the set of firms included within the initial sample, as discussed above in Section 4.1.

7.1.2 There also appears to have been an important error within the analysis approach, as now described. Within the discussion with respondents to prepare them for the stated preference tests, advice was given to them on which types of underlying cost changes were within their scope to govern their SP responses.

- All of the shippers categories A, B and C were advised (FR p.89) that they **should take account** of what would happen to the transported **products** when journey time would increase or decrease.
- The subset (A) of shippers (those who outsource **all** their transport) were advised that they **should ignore** that a longer journey time can lead to extra costs for **personnel or vehicles**.
- The carriers as well as the subsets (B) and (C) of shippers (those who outsource **only some (B) or none (C)** of their transport) were all advised that they **should take account** that a longer journey time can lead to extra costs for **personnel or vehicles** and vice versa.

7.1.3 As part of the data cleaning process (FR Table 11), the 24 shipper respondents who were working only on their own account - shippers subset (C) - were removed from the sample to be analysed. From reference to the data on company type (FR Table 7), it appears however that the 125 companies who outsource **some** but not all of their transport - shippers subset (B) – were not removed from the sample.

7.1.4 A central assumption underpinning the conversion from the estimated VTTs to the TAG values is the assumption:

- that the VTT for loaded HGVs should be the sum of the shippers plus carriers estimated VTTs;
- that the VTT for empty HGVs should be only the carriers estimated VTT.

However, the majority of shippers within the analysis sample, 125 out of 213, fall within subset (B) and these accordingly have been advised during the experiments to take account of both the product cost **and the transport** cost impacts from changes in travel times. This appears to corrupt the logic behind the proposed VTT allocation for loaded HGVs, which appears to have explicitly assumed that the shipper sample contains only subset (A).

“The SP experiments addressed this with carriers only considering transport costs and shippers (that contract out) only considering cargo costs. Therefore, empty transports were the carrier transport costs, and loaded transports were the combination of carrier transport costs and shipper cargo costs” (FR p.41).

7.1.5 Moreover as discussed in Section 5.1 above, only a small minority of both shippers and carriers confirmed in question D13 that the description of the journeys that was presented was actually clear to them. This reported lack of clarity suggests that doubts should remain on the ability of these respondents in practice within the experiments to make the required clear conceptual separation between the product cost related and the transport cost related aspects of the impacts of travel time changes.

7.1.6 In summary, the methodology underpinning the VTT and VTTR estimates for loaded vehicles that are presented in FR Table 40 is flawed due to the substantial double counting of the transport cost element, so it will greatly overestimate the correct values. It would not be a straightforward analysis task to undo this double counting because this would require separating out each of the shipper subsets (A) and (B) for individual modelling analysis but the resulting sample sizes in either case would be too small to enable successful model estimation.

7.2 THE PRODUCT RELATED VALUE OF TIME AND OF RELIABILITY

PRODUCT RELATED VTT PER TONNE

7.2.1 It is worth stepping back to reconsider from scratch the underlying concepts appropriate to the valuation of increased travel time and of reduced reliability of travel time, within the context of product related costs, rather than the context of transport related costs. This use of the VTT/R derived from product related costs based on shippers responses, is an important component of the study methodology so its appropriateness deserves scrutiny.

7.2.2 For the **travel cost element**, estimating the value of a reduction in travel time is conceptually relatively straightforward. *On average*, a reduction in travel time should lead to less time spent driving so that driver staff costs reduce. It should also allow a vehicle to make more trips or longer trips per day. Accordingly, this reduces the total number of vehicles required to transport a specified tonne kilometre total. This vehicle travel time reduction generates savings in vehicle capital and running costs. In practice, some firms in some situations may be able to fully achieve all of these transport cost savings, or to gain even larger savings, through reorganising their logistics operations. Some other firms in other practical contexts may be unable to fully achieve these savings because their vehicles and drivers are unable to be fully reassigned to other deliveries. Accordingly, in practice the VTT of transport for use in appraisal should be an estimated multiplier of the vehicle plus driver operating cost savings that result from travel time reductions.

7.2.3 In contrast, for the **product cost element**, estimating the value of a reduction in travel time is **not straightforward** from either a conceptual or an

empirical perspective. This is because it depends on a wide range of characteristics and contexts that are discussed below, which impact on the existence and the scale of the product related cost savings that a travel time saving generates.

CHARACTERISTICS OF THE PRODUCT

7.2.4 The VTT and VTTR associated with products depend on the characteristics of the product itself in a manner that varies across a detailed level of product type segmentation.

7.2.5 If a consignment has a very **high capital value**, then savings in journey time can reduce the interest costs on this capital. However, such capital based differences are unlikely to have significant implications for short-term route choice decisions in the UK because only limited time differences of at most an hour or two would arise between route options. However, in freight mode choice between domestic road, rail or coastal shipping, or between use of accompanied or unaccompanied international ferry services, the travel time differences between options may extend to some days. In mode choice, the capital value of the product may have a sufficiently large impact on the VTT to require it to be explicitly represented, for some types of more expensive products, at least.

7.2.6 Unfortunately in practice, a good, realistic representation of product capital cost impacts would require access to product type information at a high degree of segmentation differentiation to identify the vehicle load type. For example, NST:⁴⁸ - Foodstuffs is the product category with the largest traffic volume of UK road goods transported. Within it, the capital value of an HGV full load of crisps or of potatoes may be a few £1000s, while that for a full HGV load of cigarettes or of scotch may be around £1,000,000. Other foodstuffs have capital values that lie in between these extremes. Accordingly, the weighted average capital value per tonne of foodstuffs could vary greatly across the UK, depending on the main types of industries and ports in the local area.

7.2.7 Some products, particularly some foodstuffs, may be perishable when they are not in a frozen state, so they may require shorter supply chain structures than non-perishable goods. For perishable goods, accompanied ferries tend to be used rather than unaccompanied so as to minimise the number of travel days required for international trips. However, an added duration of an hour or two to a trip is unlikely to cause significant stress to the product, provided the vehicle has facilities to keep the goods chilled, while in transit.

7.2.8 Having first looked at value variations within an individual NST product type, we now switch to examine value variations between different product types - NSTs.

⁴⁸ NST – **Nomenclature** uniforme des marchandises pour les **Statistiques de Transport**

7.2.9 For the category waste (NST: 14), which is another major volume carried on road, its capital value is zero or negative. Likewise, for many construction materials (NSTs: 3 and 9) and for many bulk agricultural products (NST:1) their value per tonne is low. In contrast, for textile products (NST: 5) and equipment (NSTs: 11 and 12) their values per tonne are many multiples higher than those of the main bulks, whereas the remaining NSTs will have intermediate values.

7.2.10 The individual NST product types are not equally likely to be carried on all types of LGVs and HGVs. For example, low value bulk products and wastes have relatively low proportions of their journeys in artics, while they have a high proportion in large rigids, but with bulk construction materials also in smaller rigids. Many of the highest value goods are in artics or small rigids with few of their delivery journeys using large rigids.

7.2.11 The consequence of these observations is that the average value of cargo per tonne loaded may differ radically across different types of road goods vehicles, due to systematic differences across vehicle types in the average mix of product types that they carry. Furthermore, within a specific road vehicle type, the average value of cargo per tonne loaded may differ radically across different area of the country, and differ across the road types within an area, due to differences in the average local mix of product types produced and consumed.

7.2.12 In summary, estimating a representative overall average cargo value per journey tonne for an HGV type is not a trivial task. Moreover, there will be a large variance from this average value across the road links of the country.

LOGISTIC LEGS WITHIN SUPPLY CHAINS

7.2.13 The previous section examined how the product related VTT per tonne varies across a detailed cross-section of types of products. Now we examine how the product related VTT and VTTR per tonne for any specific product type may differ systematically across different types of journeys, depending on their stage within the supply chain. This starts with a brief background overview of logistic structures within UK supply chains.

7.2.14 Although some bulk primary products, such as ores or grains, may be transported directly without intermediate storage from where they are produced to where they are consumed or processed, this direct transfer is not the norm for most goods. Instead, foodstuffs and consumer goods usually pass through many intermediate distribution centres within the supply chain that connects the initial producer of the goods through to the final consumer.

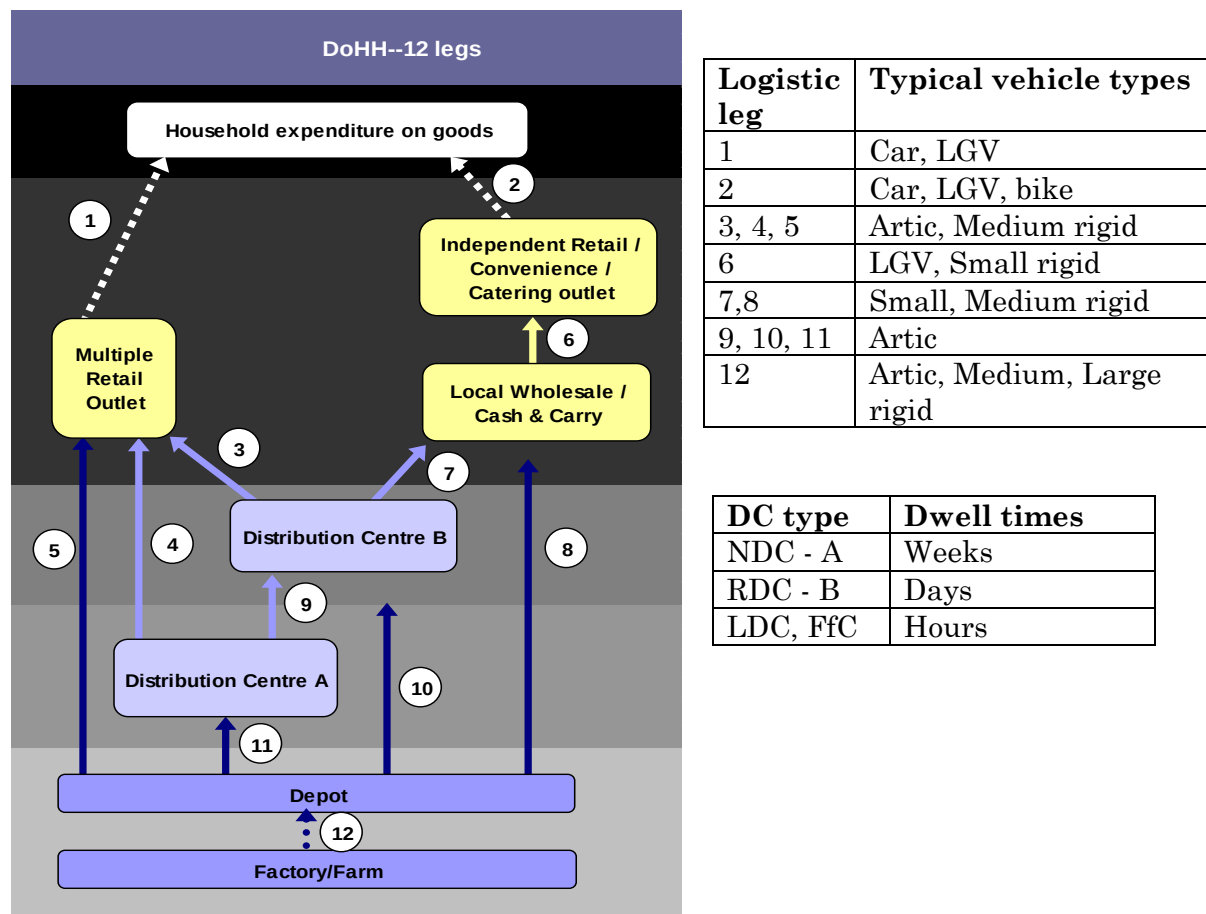


FIGURE 1 SUPPLY CHAINS AND LOGISTIC LEGS FOR CONSUMER GOODS

7.2.15 An example of some of the supply chain structures that are used to ship goods to households is illustrated in Figure 1. This chart illustrates

- Some of the different *supply chain* options that may be in use
- the intermediate *distribution centres* (warehouses) at which goods are stored
- the individual *logistic legs* travelled between each intermediate storage stage within a supply chain.

7.2.16 In general, when consumer goods leave the producer's depot they are likely to travel to a large central National Distribution Centre (NDC - A), being shipped as a large consignment within a large artic. Typically many of the items within this consignment may be stored at the NDC for some weeks before they have gradually all been shipped out to different regional locations, again mainly within large artic. The goods in a consignment will spend a few days at a Regional Distribution Centre (RDC - B) before gradually being shipped within an artic or a medium rigid HGV as smaller consignments to either a large supermarket or retail outlet or else to a Local Distribution Centre (LDC) or a local parcel delivery hub. From the LDC, goods may be delivered to or collected by smaller local retailers or caterers within LGVs or small rigids. Alternatively, items may be delivered from a local parcel delivery hub directly to households or firms as parcel deliveries from vans or cargo bikes doing a local round. In recent years via online ordering procedures, groceries in large

supermarkets are increasingly being delivered in LGVs direct to consumers, rather than being collected by retail customers in their own cars.

7.2.17 The aim underpinning the evolving design of these supply chain structures within firms is to minimise the overall cost of their goods distribution system (the costs of transport, warehousing, capital, etc.), while meeting the consumption needs of their customers. The most cost-effective way of avoiding temporary stock-outs is through carrying substantial stock levels in a central NDC from where items can then be distributed when and where they are needed, while minimising the requirement to carry large stocks at a regional and especially at a local level. This is why goods will typically be stored: for weeks in NDCs; for a few days in RDCs; and often for just a few hours in local parcel delivery hubs. The average consignment size required at the delivery end of a logistic leg typically decreases when progressing through the individual stages of the supply chain from the producer through to the consumer. This is why the typical vehicle size used gradually decreases through the sequence of logistic legs from the producer to the consumer.

7.2.18 The supply chain system just described enables some of the complexities and apparent contradictions in VTT / VTTR valuation to be understood better. Because of their large stocks and long dwell times, a delay or a longer journey time for an artic feeding an NDC may often have few wider implications beyond its impact on the transport cost. In contrast, delays in deliveries to or from parcel delivery hubs may generate many irate customers. Because artic (OGV2) will supply most of the deliveries to NDCs, whereas small rigids or LGVs will supply most of the final-leg deliveries to consumers, for a given type of product the latter will tend to have much higher product related costs of delays than most artic. Due to the logistics behaviour illustrated above, a smaller HGV may have a higher VTT than a larger HGV, which might explain some of the “apparently odd” results generated in the model estimations.

7.2.19 More generally, if road speeds decrease on average across the UK, then a greater number of LDCs within a revised UK supply chain structure would be needed to serve the UK consumers in a timely fashion. This implies that the overall cost of supply chains would increase accordingly, over and above the added costs of the transport itself. This is a long-term cost that should be captured within appraisal procedures, even though it may not have a direct impact on short-term route choice behaviour.

7.3 Summary of implications for VTT/R estimation

7.3.1 This Section summarises the findings from the material above to support its review of their implications for the estimation and usage of VTT and VTTR within TAG.

7.3.2 The difference in the likely level of VTT, and especially of VTTR, for different types of trips to port has been discussed above in Section 3.2. It explains that an artic on its way to initiate an accompanied crossing on a ferry that is leaving soon,

would have a much higher VTTR than would HGVs on unaccompanied ferry services or HGVs accessing ports for other types of deliveries.

7.3.3 Summarising the discussions in Section 7.2 for the product related VTT and VTTR per tonne carried, indicates the following characteristics.

- a) Their values will vary substantially across different product types, related mainly to the product unit cost and whether perishable or not, noting that for many low value bulk commodity types these values may be minimal.
- b) For any given product type, their values typically will **increase** significantly along the sequence of logistic legs from producer to consumer.
- c) For any given product type, because the average goods vehicle size tends to **decrease** along the sequence of logistic legs from producer to consumer, their VTT and VTTR per tonne carried may accordingly increase as vehicle size decreases.
- d) The average mix of product types carried is likely to vary widely across vehicle types, so their average product related VTT and VTTR per tonne carried will accordingly vary widely.
- e) For a specific vehicle type, the average mix of product types carried may vary considerably between areas and between road types, so their average product related VTT and VTTR per tonne carried may also vary considerably.

7.3.4 These findings highlight: that the estimation of product related VTT and VTTR per tonne carried, is highly context specific; and that for any given vehicle type their values will vary widely and will be influenced by a wide range of inter-correlated explanatory variables. Accordingly, the estimation procedure would require a much larger sample than that which had been envisaged originally for this study, to ensure it could provide a sound model estimation of their values in a form that successfully tested the influence of all of these inter-correlated effects. This would not be a simple modelling exercise and it is unlikely to be feasible to implement it through using the sample size available within the current study.

7.3.5 On the other hand, it appears likely from the logic of the discussion above that the product related VTT per tonne carried would be rather small for a substantial proportion of all HGV journeys, particularly those in the larger HGVs, since these HGVs are less likely to be being used for the later stages in supply chains. However, it would be wise to empirically test that in practice this deduction is observed to be well founded.

7.4 Usage in TAG of the empirical results from the study

7.4.1 The discussion here aims draws the findings together

- to assess the usability within TAG of the empirical results from the existing experiments
- to ascertain whether some reanalysis of the SP survey data might be informative.

Guidance on how future studies might best be oriented is then provided in Section 8.

7.4.2 The examination within this peer review of issues that have arisen during the course of the sample development, of the SP experiments and of the model

estimation stages, indicates that not all of the original aims of this study can be achieved. The potential usability within TAG of the current model estimates for each of the six main segments is summarised in [Table 8](#).

[Table 8 Usability in TAG of estimated results by segment](#)

Segment		OK?	Comment
Shipper	LGV	No	The inclusion of Shippers (B) in the shippers analysis corrupts the conceptual approach, while their exclusion leaves too small a shippers sample (See Section 7.1)
	OGV1		
	OGV2		
Carrier	LGV	Perhaps	Problems with BVTT options (Section 5.2). Sample very small and not-representative (50% port related)
Carrier	OGV1	Maybe	Sample not-representative (only large hauliers)
	OGV2		

7.4.3 However, this unavailability of acceptable results for the shippers segment may not be such a major issue, because the discussion in the previous Section has cast doubt on the scale and universality of a large product related addition of VTT and VTTR by comparison with the established important transport related component. Instead, **if** this product related component is relatively small in reality for many (but certainly not all) road freight journeys, then the current inability to measure it becomes less problematic.

7.4.4 It is important for TAG to consider the various modelling or appraisal contexts in which the VTT and VTTR are used. The appropriate VTT definition and value to utilise, may differ between the different model choice stages and for appraisal, as summarised in [Table 9](#), related to the likely relative importance of the inclusion of the product related cost element for that stage.

[Table 9 Relative scale of product related VTT and VTTR impacts, by model stage](#)

Stage	VTT/R
Road route choice	Small for many but not all GVs
Mode choice	Medium
Appraisal	Large

7.4.5 It is expected that in general the decision on road route choice will be made by the transport operator or driver for whom the commodity type of the contents of the vehicle may not be a material issue with respect to their short-term route choice. In any case, the time savings available from switches in route would be relatively small. Logistic stage may be a more important determinant of VTT/R levels, because final deliveries to consumers may be more time sensitive, and these are often in LGVs and smaller rigids.

7.4.6 For mode choice because of the systematic potentially large time and reliability differences between road or accompanied RORO versus other modal options, the product related VTT and VTTR impacts on decision making are likely to be more substantial.

7.4.7 For appraisal, the empirical impact of both the transport related and of the product related VTT/R should be increased to take account of the longer term logistic benefits. An example can illustrate why. If there was a future policy to increase the speed limiters for HGVs from 90 to 100 kph and to increase the speed limit for LGVs to match that for cars, this would certainly lead to transport related cost savings through the reductions in drivers' hours and in the number of vehicles required, although with an increase in the fuel consumption rate. However, it would also facilitate reorganising the overall logistics structure for firms to reduce their costs by requiring fewer RDCs and especially fewer LDCs to avail of the larger catchment areas that could be replenished in a timely fashion, based on these faster road speeds. These longer term logistic cost savings (TAG – Wider economic impacts) from freight transport time reductions are relevant to appraisal, though not to the choice modelling stages.

7.4.8 Having accepted that the SP experiments for OGV carriers might have potential to provide useful results for TAG, the question arises as to whether some further re-analysis could be helpful. Adjustments to explore could include the following.

- a) For each vehicle type separately, reweight down its sample entries for the segment of journeys to port, in order to correctly represent their small proportion across the GB population of journeys, based on CSRG proportions. This weighted regression should generate different answers to the current approach of including a port specific constant among the explanatory variables. It could also be used in reweighting the final results to expand to the GB population.
- b) Undo the forcing of the OGV2 VTT to be no smaller than the OGV1 VTT. Investigate further whether there may be some rational explanation underpinning the original significant variables that were dropped due to their assumed “counterintuitive result” (FR, p.39).
- c) Investigate and understand better the influences underpinning the individual exclusion criteria to see whether their application or non-application in reducing the analysis sample should be adjusted.

7.4.9 The existing VTT comparison in FR Table 43 shows that the carrier based VTTs for LGVs and OGVs of this study are 4.1 and 3.5 times higher, respectively, than those currently within TAG. The TAG derivations are based solely on estimated driver plus vehicle operating costs per hour.

7.4.10 Given the many caveats discussed above, including the lack of representativity of the samples and the apparently high level of lack of understanding of context among survey respondents, it appears prudent to **not replace** the current TAG approach and values until greater confidence can be

reached about the solidity of the foundations of these new methods and of their empirical results.

7.4.11 In the next Section guidance is provided on various ways in which future SP studies could be improved. It may also be appropriate to explore VTT estimation through revealed preference methods as discussed below in Section 8.4, together with some potential examples of practical applications.

7.5 RECOMMENDATIONS FOR OTHER TAG ADJUSTMENTS

7.5.1 In the course of examining the methods for introducing the findings from the study into TAG, some recommendations for other potential improvements to TAG itself have emerged from the review.

7.5.2 The first improvement relates to the proposed manner for converting the LGV VTT into a form suitable for use in TAG. FR p.42 refers to “the 88:12 LGV freight: non-freight ratio in Tab A1.3.4 of the TAG Data”. The legitimacy of this TAG ratio should be checked by the Department. It does not appear to match to that in any of the various past DfT surveys of LGV activity. For example, the 2020 DfT van survey⁹ indicates that 24% of van mileage has delivery/collection of goods as the primary usage, which differs greatly from the TAG assumed 88%! It is important to have a clear distinction within TAG, as well as in the underlying transport models that make use of TAG, between: the minority of LGV journeys that are primarily for freight purpose – defined as the delivery/collection of goods; and the majority of LGV journeys that are for other non-freight purposes. Most van journeys for commuting purpose or for in course of business trips by tradesman other than professional van delivery drivers, will indeed also be carrying goods – that is why vans rather than cars are used. However, the carriage of the tools of the building trade and those of other manual trades should not be confused with the goods delivery/ collection freight trip purpose. Appropriate VTT and VTTR parameters should be allocated in TAG to each distinct LGV journey purpose: freight; commute; in course of business; leisure / other.

7.5.3 The second improvement would resolve a current apparent incoherence in representation within TAG between:

- Unit M3.1 Highway Assignment Modelling, which states: “The value of time given in [TAG unit A1.3](#) for HGVs *relates to the driver’s time* and does not take account of the influence of owners on the routeing of these vehicles. On these grounds, it may be considered to be more appropriate to use a value of time around twice the [TAG unit A1.3](#) values.” (Tag Unit M3.1, para. 2.8.8, italics added)
- Unit A1.3 User and Provider Impacts, which states that “the elements making up non-fuel vehicle operating costs include oil, tyres, maintenance, depreciation and *vehicle capital saving* (only for vehicles in working time). The non-fuel elements of VOC are combined in a formula of the form

⁹ Table VAN0211 <https://www.gov.uk/government/statistics/van-statistics-2019-to-2020>

$$C = a1 + b1/V$$

where;

C = cost in pence per kilometre travelled;

V = average link speed in kilometres per hour;

a1 is a parameter for distance related costs defined for each vehicle category;
and

b1 is a parameter for *vehicle capital saving* defined for each vehicle category
(this parameter is only relevant to working vehicles).” (Tag Unit A1.3,
para.5.1.10, italics added)

7.5.4 For all vehicle types within the highway assignment modelling stage, the generalised cost that is utilised is in time units rather than in cost units.

Accordingly, this distance-based non-fuel cost term C, is converted to be in time units through dividing it by the value of time (Tag Unit M3.1, para. 2.8.1).

7.5.5 Rather than arbitrarily doubling the HGV driver’s value of time, as is presently recommended in TAG Unit M3.1, it would instead be more logical to add the vehicle capital saving element to the HGV driver’s VOT to create a more appropriate HGV value of time for use throughout. Furthermore, it would be preferable to extend the vehicle capital cost saving component by appending to it all other time dependent vehicle costs, such as vehicle insurance, Vehicle Excise Duty, Road User Levy, etc. In this way, a more logical justification can be adopted to achieve the original requirement of increasing the size of the parameter value.

7.5.6 A third recommendation is a change in emphasis to the guidance that relates to the TAG Unit M3.1 discussion on route choice calibration for HGVs. Relating to highway assignment calculations, TAG currently states that

“The operating cost formulae given in TAG unit A1.3 relate vehicle operating costs to speed of travel, by vehicle type. It is not considered necessary to use the speeds of traffic on each link individually” (TAG Unit M3.1).

For the reasons now explained, while this may be an acceptable simplification for cars, it should not be adopted for HGVs and it is not ideal for LGVs delivering goods.

7.5.7 The TAG section on “Route Choice Calibration for HGVs” states.

“It is often the case that the routes based on generalised costs given in TAG for heavy goods vehicles do not appear to take full account of the attractiveness of motorways and trunk roads and the unattractiveness of local roads for these vehicles.” (Tag Unit M3.1, para. 7.2.2)

7.5.8 The current advice on resolving this issue (para. 7.2.3) is to replace:

- a) the existing widespread practice that adopts a common average road speed throughout the study area when calculating HGV route costs; by
- b) instead adopting an alternative link-based calculation of HGV generalised costs that would reflect the real world speed differences by link (or at a minimum by link type).

7.5.9 This alternative should be given a much stronger emphasis in TAG for HGVs because it avoids¹⁰ the current systematic overestimation of the time-based HGV non-fuel costs on fast motorways and trunk roads and their underestimation on slower, single carriageway roads, which is a major reason causing HGV assignment patterns to be problematic.

7.5.10 Moreover, the HGV speeds used throughout these link-based calculations should be consistent with the HGV speed limits specific to the road type for each link, not simply adopting the modelled car speeds. Assignment model performance would be systematically improved by stressing the importance of these refinements to HGV cost estimation within assignment procedures.

7.5.11 This link-cost problem does not arise for passenger vehicle assignments (other than marginally for the trip purpose: in course of business) because as stated above, the time-based cost term “b1” is used only for “working vehicles”.

8 Recommendations within future studies

8.1 Improving the segmentation within the sample

8.1.1 There are various ways in which the sample design could be improved if future VTT studies are initiated.

8.1.2 Firstly, it is debateable whether there was need to explicitly distinguish in the experiment those trips to ports for the OGV1 and LGV segments. The proportion of all LGV¹¹ trips or of all OGV1 trips that actually are to ports or to airports is likely to be minimal within most of Great Britain, except in a few areas adjacent to the Isle of Wight or the Scottish Islands. [Table 10](#) illustrates the percentage of journeys that are to or from intermodal terminals for each type of HGV. Although artics have over 6% of their journeys with an end at an intermodal terminal, for the various rigid classes this percentage is below 2%.

¹⁰ The scale of the errors can be demonstrated through contrasting the calculation of the non-fuel cost function C in the case where the speed divisor V is identical, irrespective of the real link speed, versus the correct result produced by adopting the actual speed on each individual link, or at a minimum through adopting the average speed for each specific road class.

¹¹ According to the DfT 2019-20 Van Survey Table VAN0302, only 0.5% of vans whose primary usage was for the collection/delivery of goods are typically used for international deliveries. Most “vans” typically used for international travel had their primary usage for recreational/leisure and holidays, so they were more likely to be motorhome type vans than be delivery vans.

Table 10 Percentage of journeys to / from intermodal terminals by type of HGV, GB

Intermodal facility	Rigid				Articulated	Total
	>3.5t to 7.5t	>7.5t to 17t	>17t to 25t	>25t		
Airport	0.5%	1.1%	0.7%	0.0%	0.3%	0.3%
Rail Siding/Terminal	0.0%	0.0%	0.3%	0.8%	1.0%	0.7%
Shipping Docks	0.0%	0.8%	0.8%	0.5%	5.1%	2.8%
Not Inter-modal	99.5%	98.0%	98.3%	98.6%	93.6%	96.2%

Source: CSRG, 2022

8.1.3 Accordingly within a limited overall budget the sample quota devoted to the 1% of OGV1 trips or to the 0.5% of LGV trips travelling to/from intermodal terminals is unlikely to provide good value. It appears likely that removing the requirement for half of all respondent firms to be providers of shipments to ports / airports, would facilitate recruiting a much greater proportion of respondents in small to medium firms within the sample. This is an important requirement for achieving a representative sample profile.

8.1.4 Secondly, there is a lack within the sample design as well as within the subsequent experiment design, of any distinction within the trips to ports between:

- those accompanied HGVs that almost immediately expect to board a ferry;
- versus those HGVs just depositing trailers, containers or bulk commodities at the port, which are then scheduled for onward shipment over the coming days.

As explained in Section 3.2 above, the VTTs and VTTRs are likely to differ strongly between these two categories of HGVs which is why they would benefit from being in separate segments.

8.1.5 Thirdly, within the OGV2 category the proportion of port-bound journeys is likely to be heavily biased towards articulated rather than rigid HGVs, except perhaps for the subset of ports focussing on bulk products. It is unhelpful to use a single OGV2 segment which combines articulated and rigid vehicles with 4 or more axles, rather than adopting instead a distinction within it between large rigid and articulated HGVs. The CSRG together with other DfT statistics tables indicate that these two vehicle types carry very different mixes: of commodity types; of values per load; and use very different patterns of road types during very different patterns of time of day and day of week. For example, the proportion of large rigid vehicles on accompanied mode through ports will be tiny. In contrast, the proportion of rigid vehicles delivering bulk goods to silos at ports will be significant.

8.1.6 The precise definitional distinction to adopt between the small and large rigid HGV classes, requires further investigation and analysis, because there may be good analytical reasons to not automatically match small rigids across to the traditional OGV1 category. In fact, it may be preferable to distinguish all four standard rigid classes, because they tend to differ in their usage within supply chains and across product types, as discussed above.

8.1.7 Segment the journeys to port only for those artics travelling to board ferries (accompanied RoRo). Journeys to ports or airports for small rigids and for LGVs will be relevant only to a very small proportion of their respective markets, so they are of low importance in the context of Great Britain as a whole. Journeys to ports for unaccompanied artics, and for containers or bulk cargos do not generally have tight deadlines for delivery, as the goods typically are stored on quayside for some time prior to being loaded onto vessels. For the remaining vehicle types, other than the artics to ports, some refinements to the questions B15 to B20 in the current survey, which relate to time windows and delay costs, should be able to provide an alternative foundation for investigating the impact of journey deadline pressures on VTT and VTTR estimates.

8.1.8 In summary, this proposal reduces the number of modelled vehicle segments from 6 classes down into 5 classes: LGV; small rigid, large rigid, artic to board ferry; other artic journeys and so it would be cost effective. This approach would increase within class homogeneity of behaviour, which should improve the performance of the modelling stage. Populating these 5 segments with relevant and representative respondents during the sample assembly stage should be less challenging than occurred within the current study. There would be benefits from further increasing the segmentation of rigids from 2 to 3 or even 4 size classes.

8.2 SAMPLE SIZE AND SURVEY DESIGN IMPROVEMENTS

8.2.1 A study based on a substantially larger initial sample of respondents, classified into a somewhat different segmentation would have had the potential to alleviate a number of the issues and uncertainties that arose in the current study, for reasons already discussed in earlier sections.

8.2.2 The inherent high level of heterogeneity of freight transport response behaviour implies that a relatively large sample of respondents is required in order to adequately represent those responses particular to specific homogenous segments within the overall population of road freight firms in Great Britain. Moreover the initial sample needs to be large enough to offset the likelihood that some respondents will later need to be excluded from the model estimation step for various reasons that may be difficult to predict. It is crucial that an adequately sized and representative sample of fully responding SP experiments can be achieved, for which there is confidence that each person is responding within a context that is familiar to them individually and that is representative of the costs, times and other transport characteristics within a journey context that they regularly confront.

8.2.3 The set of five exclusion criteria adopted in the current study should be maintained, though with an adjustment now described. Distinguish between the SP1 VTT question set and SP2 VTTR question set when querying whether respondents have fully understood the context and questions in order to populate the exclusion criteria (C2) to (C4). This requires asking the appropriately adjusted questions D12 to D15 at the end of the VTT Section C, in addition to where they currently are asked at the end of the VTTR section D. Then initially apply the

appropriate exclusion criteria relevant to the VTT model and subsequently apply those relevant to the VTT plus VTTR model. This should ensure that any lack of clarity regarding the travel uncertainty VTTR experiments, would not indirectly reduce the sample size available for the earlier VTT analysis.

8.2.4 It is envisaged that this future adoption: of a more representative sample of respondents; of more homogenous individual segments that are less unbalanced in population size; and of more discriminating exclusion criteria questions, should in combination strongly reduce the proportion of the initial survey sample that would be lost during the exclusion criteria stage. Nevertheless, the initial sample size adopted needs to be sufficiently large to allow all of the exclusion criteria to be applied¹² while still retaining a large enough resulting analysis sample to enable flexible model testing to be achievable at the detailed individual segment level.

8.2.5 In summary, start with a sufficiently large sample to ensure that the exclusion criteria do not subsequently shrink down the sample sizes for any individual segment that fall below the quota required to ensure an adequate level of explanatory power and precision within its estimated models. The quota selected need to be large enough so that the model has the power to identify the significant external variables that impact on VTT and VTTR levels, noting that the discussion above in Section 7.2 indicates that many of these explanatory variables are likely to be inter-correlated which will require some sophistication within the model development and testing. These external variables can then be used to offset such biases as may have emerged within the sample, so as to ensure that the finalised VTTs and VTTRs when expanded from the sample to the population are then representative of Great Britain as a whole. These requirements imply that a substantially larger initial sample than the current 600 would be needed in order to provide a safer analysis environment that would increase the likelihood of success and of sufficient precision within the overall estimation procedure.

8.3 EXPERIMENTAL DESIGN IMPROVEMENTS

8.3.1 This Section proposes three adjustments to the approach to the SP experiments in order to improve the model estimates of VTT and VTTR.

8.3.2 Firstly, as discussed above in Section 5.1, the indirect evidence from the high failure rate on exclusion criterion (C4) suggests that many of the respondents may not have had a clear understanding of the experiment. A critical assumption underlying the evidence gathering approach within this study is that it is feasible, while assessing their value of journey time or reliability savings within their SP experiment:

¹² A major source of uncertainty for the estimated VTT and VTTR within the current study, arose because of the inability to apply the full set of exclusion criteria without reducing the sample size down to a size that was too small to enable effective model estimation.

- for each shipper respondent with no in-house transport provision (i.e. shipper subset A) to focus **only** on “what would happen to the transported **products** when journey time would increase or decrease” (FR p.89);
- for each carrier respondent to focus **only** on the savings or increases in vehicle and driver **costs of transport** associated with changes in journey times.

8.3.3 This reviewer has some doubts (but no explicit hard evidence to support such doubts) about the universality of the ability of the sampled respondents to make this refined, detached judgement, while responding within a busy office environment and while being subject to time pressures to get the survey experiments finished quickly so as to allow normal commercial tasks to be reinitiated. The recommendation in paragraph 8.2.3 above for more discriminating questions regarding the level of clarity of the experiment perceived by respondents, should help in explaining whether their uncertainties are focussed mainly on the SP2 - VTTR or else relate to both SP1 - VTT and SP2 - VTTR.

8.3.4 One potential future way of testing whether this evidence gathering approach within the experiments has been successful in practice would be to further increase the segmentation adopted at the model estimation stage to now explicitly distinguish each of the subsets (A), (B) and (C) of shippers (respectively those with none, some or all of their transport provision being in-house, rather than being from external hire-and-reward hauliers). So long as the resulting sample size quotas are sufficiently large for every segment, this adjustment would enable the consistency of the study approach to be confirmed, based on there being no significant difference in estimated values between: the sum of the VTT of Carriers plus the VTT of Shippers (A); versus the VTT of Shippers (C). This equivalence arises because this Shippers (C) class is required within the experiment to consider the impact on both the product delivery and on the vehicle and driver costs. Furthermore, the VTT of Shippers (B) should fall within the range between that of Shippers (A) and of Shippers (C), in a pattern that is influenced by the proportion of in-house transport provision of each Shipper (B) respondents. This methodology consistency test would need to be applied individually in turn for each vehicle type segment.

8.3.5 This requirement to increase the segmentation of shippers from 1 to 3 classes would lead to an overall doubling in the total sample size that is needed. If this consistency test is successful, it would provide solid evidence of the suitability of this underlying methodology which sums a transport related VTT plus a product related VTT in order to produce a total VTT suitable for use in assessment. Alternatively, if the methodology consistency test does not hold, then at least it should be feasible to examine across the four distinct segments of types of firms, their mutual patterns of estimates of VTT, VTTR and of significant external explanatory variables. This examination should facilitate a better understanding of how to determine VTT and VTTR values suitable for use in assessment and should help to explore further whether the carriers VTT is equivalent to that for an empty vehicle.

8.3.6 Secondly, from a methodological viewpoint it would always be preferable to initially estimate a separate individual model for each individual segment, rather than to pool the data within a combined segments model within which individual segment-specific constants are used to account for each of the individual segments. This initial individual model approach will increase the size of the sample quota needed but it should provide more appropriate estimates of VTT and VTTR for each individual segment. Provided that the resulting structure of their significant independent variables is similar across the individual models for a subset of segments, then a combined segments model would be suitable to be applied for this specific subset of segments. Because of the resulting larger pooled sample, this combined segments' model should then provide a higher level of precision for each of its estimated parameter values.

8.3.7 Thirdly, ensure that the range of BVTT values presented within each experiment is realistic for every individual segment and that it avoids the erratic patterns across segments that were detrimental to the current study (Section 5.2). For carriers and for those shippers with own-account fleets, the determination of the BVTT range should be switched from a per tonne basis to instead be on a per vehicle basis. In order to avoid biasing the results that are generated by the methodology confirmation test proposed above in paragraph 8.3.4, the BVTT range for a given vehicle type should be similar across three of the type of firm segments, comprising carriers plus the two shipper types that have some or all of their transport within their own-account fleet.

8.4 POTENTIAL FOR REVEALED PREFERENCE RESEARCH

8.4.1 Revealed Preference (RP) based studies could provide an alternative approach to the estimation of VTT and they could be used to complement or replace the current SP based study. The following examples illustrate contexts in the UK that could be suitable to support such RP studies.

8.4.2 Firstly, for those firms importing or exporting goods by road ferry between the UK and the EU, there is a choice between: the use of accompanied HGVs (fast and expensive); rather than unaccompanied trailers (slower and cheaper). An informative study would examine for a variety of O-D pairs which types of goods under which types of conditions use each type of service. How does this split differ by commodity type, by value per tonne, and by degree of perishability of the goods? There is scope also to investigate how and why the proportion of use of unaccompanied trailers has increased strongly during the period in which accompanied HGVs are experiencing the increased vehicle costs that arise from the border delays that are associated with Brexit.

8.4.3 This type of ferry crossing based RP study would capture the wider benefits from large savings in travel time and so would be relevant for use for parameters in mode choice models.

8.4.4 Secondly, the use or avoidance by goods vehicles of UK toll roads or bridges could be studied. The sites which experience expensive tolls and for which there are substantial potential goods vehicle movements, include

- Dartford crossing £6 / £3 for HGV>2 axles / LGV & HGV 2 axle
- M6 Toll route - £15.90 / £15.30 for HGV>2 axles / LGV & HGV 2 axle – with a lower charge for partial distance usage of the route
- Humber bridge toll - £12 / £4 / £1.50 for HGV>2 axles / HGV 2 axle <7.5 tonne / LGV.

These toll levels are the 2024 standard vehicle charges before any discounts are applied.

8.4.5 The revealed preference behaviour at these toll options could be measured in two ways. Firstly, remote studies could use number plate matching to check the route travel time differences and the route choice percentage splits, across different times of day/week by vehicle type. The vehicle number plates would need to be captured both before the route split at the start, as well as after the routes rejoin together past the end of the toll route. The number plates would also need to be captured at one suitable location along each of the toll and of the alternative route, so as to determine which route option had been selected for each surveyed vehicle. This approach would provide data suitable to estimate VTT in a form segmented by LGV (but not by LGV journey purpose) and by HGV class, segmented by vehicle type and size. By virtue of the impacts of uncertain congestion delays expected at certain times of the week, it may also produce estimates of VTTR for each of the goods vehicle type segments.

8.4.6 Because the choice of the M6 Toll route versus the M6 is relevant mainly to relatively long distance inter-urban traffic, it may not capture a representative cross-section of goods vehicle traffic as a whole, especially for some types of rigid HGVs due to their short average trip lengths compared with those of artics. This uneven coverage would need to be resolved in some way as part of the analysis of results to ensure that the final expanded results are representative of the UK HGV traffic as a whole.

8.4.7 It is necessary within the RP to separate out LGV delivery trips from the other LGV trip purposes that are not of interest to this freight study. Because a specific subset of high capacity van marques tends to be widely used within the delivery sector, it may be feasible to focus on these specific van marques or else on those vans with a livery denoting their delivery industry sector, when carrying out the remote LGV route choice survey for the RP.

8.4.8 Alternatively, a representative cross-section of the firms that are likely to regularly face toll route choice options for their fleet could undergo a structured interview that examines the conditions under which these firms have made use of the tolled and of the non-tolled options during the period under study.

8.4.9 This type of toll-based RP study would capture the immediate benefits from relatively small savings in travel time and so would be relevant for use for route choice parameters in assignment models.

8.4.10 In summary, the inherent level of uncertainty over the appropriateness of the VTT/R results could potentially be reduced through the examination of the revealed preferences, rather than the stated preferences, of a representative cross-section of road freight firms. However, more detailed prior investigations would first be needed to refine the details and to confirm the likelihood of success from the estimation of revealed preference based estimates of VTT within studies of the types described above within this Section.

9 References

AECOM, Arup, Significance, ITS Leeds (2023) Freight value of time and value of reliability: Final Report. Research study for National Highways and the Department for Transport.