# Response to “Peer review of Freight Value of Time and Value of Reliability”

Thijs Dekker, Institute for Transport Studies, University of Leeds
Gerard de Jong, Significance

October 2025

## Introduction

ITS Leeds, Arup, AECOM and Significance carried out a study on road freight values of transport time (VTT) and values of transport time reliability (VTTR) for National Highways and the Department for Transport (DfT). DfT then commissioned independent consultant Ian Williams to review the final report. At the time of receiving the review report in February 2024, Gerard de Jong (Significance) and Thijs Dekker (ITS Leeds) provided a point-by-point response to Ian Williams’ review in the form of annotations in the review document and shared those with the DfT.

This note synthesises our response to the Ian Williams review and brings together the main messages captured by the referred point-by-point responses. In developing the original annotations support was received from Arup and AECOM. This note, however, reflects our personal opinions and not necessarily those of Arup and AECOM.

Before responding to the review, we would like to thank Ian Williams for the thorough examination of the original report and providing constructive feedback on the work conducted. Especially the final chapter of the review provides useful recommendations for the design of potential future studies in this area. These recommendations, however, fall beyond the scope of the study which was subject of the review.

Overall, we are pleased to read that the review considers the work of good standard. Nevertheless, Ian Williams makes valid observations regarding the size and quality of the sample. In this document, we provide additional clarifications and, especially in relation to the need to collect a representative sample, present a rebuttal.

## Sampling – segmentation and representativeness

We wholeheartedly agree with Ian Williams that the freight sector can be characterised as a heterogeneous industry comprising many different operators largely varying in size, the types of services they provide, and how they provide these.

One of the objectives of this study has indeed been to understand heterogeneity in preferences regarding the importance of travel time savings (and reliability). To identify heterogeneity in preferences across different population segments using discrete choice models, however, does not require a fully representative sample. Instead, what is

required is a sufficiently large sample within each population segment (or cell) of interest to statistically identify the presence of such preference heterogeneity across segments. In other words, the main requirement for our sampling strategy is that we have enough observations for the estimation of separate coefficients (e.g. VTT), not representativity.

Representativity in sampling also features in the recent scoping study for a new passenger VTT study for the DfT (ITS et al. 2025). There it is stated that "*To obtain a good coverage of specific variables of interest this may mean that we may wish to oversample short- or long-distance trips or specific population cohorts. Ultimately, this will enable better identification of the parameters of interest than when using a fully representative sample.*"

In the maximum likelihood estimation of discrete choice models, the part of the likelihood function that matters for the coefficients that are to be estimated to derive the VTT is called the kernel. The sampling fractions are not part of the kernel. Consequently, one can estimate the coefficients consistently on a sample that is exogenously stratified (e.g. oversampling of large firms) or, except for the alternative-specific constants, even endogenously stratified (e.g. a mode choice model estimated on a sample that oversamples public transport). We refer to Chapter 8 in Ben-Akiva and Lerman (1985) for more details. For this study, overall representativity (or alternative put obtaining a truly random sample) was therefore not an objective.

The explicit reference by Ian Williams with respect to the 'to port' sample acts as a case in point. The oversampling of this segment does not influence our ability to estimate consistent estimates for other parts of the population.

Where the modelling exercises are targeting the identification of preference heterogeneity and thereby may benefit from non-representative sampling strategies, the implementation of the models does require consideration of representativity. For example, when deriving equity-weighted VTT measures the implemented population weights should consider the frequency a given population segment occurs. In the 2014/15 passenger VTT study, for example, representativity is ensured through the sampling enumeration process making use of the sampling weights comprised in the National Travel Survey (NTS) over the 2010-2012 period. In the Freight VTT study a similar approach was adopted using the Continuous Survey of Road Goods Transport (CSRGT).

For this study, we set out with a target segmentation strategy of the VTT for different types of vehicles (LGV/OGV1/OGV2), types of trips (To port / Other trips), and types of operators (carriers/shippers). This segmentation strategy was agreed with the client group. Indeed, additional segmentations may be relevant for such a heterogeneous sector, but that would on the one hand require a much larger sample with significant cost implications. On the other hand, the ability of transport models to handle all the different levels of variation is limited and accordingly the issue of proportionality needs considering.

The adopted sampling strategy and size is not too distinct, and in some regard even larger, relative to other international freight VTT studies. In this way, our approach is typical of the approach that other government agencies have taken. This has not been recognised in the review. Moreover, our target of 50 observations per cell (or segment) does consider that some responses will be discarded, and simple choice models can be estimated across samples where up to 20% of observations can be removed, obviously this is conditional on the quality of the responses provided.

Overall, the chosen level of segmentation is always a subject of debate in any national VTT study regardless of whether this applies to passenger or freight transport. Accordingly, we highly recommend the department in future studies to review its envisaged degree of segmentation and potentially increase the number of population segments when this is proportional, and feasible in terms of sampling strategies. Particularly the latter consideration is important because achieving the present sample sizes turned out being a non-trivial exercise.

## Sampling – response rates

In chapter 4 of the review, Ian Williams makes a valid observation that the response rate of below 10% to our survey has been low, and that the elicited values of travel time and reliability for respondents might be quite different to those who did not respond.

Response rates in freight research are challenging, and the response rates observed in other national freight VTT studies (e.g. Norway) have sometimes been even lower and the obtained results have still been used in practice. This indeed poses the challenge of bias in the collected sample. The only way to establish the presence of such bias is by carrying out a non-response survey. Whilst this might be useful, similar challenges in recruitment will be experienced and the results are often inconclusive due to limited sample sizes. A non-response survey was not included in the original study design and accordingly we cannot comment on the size of any potential bias in the results.

A further observation is made that the obtained sample under-represents small firms. Whilst we agree that a limited number of small firms are included in the sample, the presented comparison is not entirely like for like since our collected sample covers both shippers and carriers, whereas the review only focuses on the high representation of small firms in the UK's haulage sector. Again, the aim of the study was not to obtain a representative sample of the UK population of hauliers. Indeed, if firm size is deemed an important level of segmentation in future studies, the sampling strategy should take this into account.

Section 4.2 of the review zooms in on the study not achieving the quota totals for several identified segments. As mentioned before; in setting these quota targets some leeway has been built in accounting for the fact that not all collected data would eventually be suitable for use. This, however, does not take away that working with limited sample sizes

limits our ability to find significant differences in the metrics (VTT and VTTR) of interest. It goes without saying that ideally, we would have worked with larger samples but unfortunately collecting the present sample already proved challenging.

Ian Williams also correctly observes that the reported average costs per journey hour from the sample has indicated cost relativities between vehicle types of OGV1>LGV>OGV2. This does not match the expected sequence of OGV2>OGV1>LGV based on known average vehicle running costs by vehicle type from TAG. This is a good point and indeed contrary to expectations. A possible explanation could be that carriers ask a premium for a short time between order and delivery, which usually coincides with the use of smaller vehicles. There can be many other explanations, but we have no means of verifying these. Nevertheless, the finding on the cost relativities in the reference trips originate directly from the sample and does not automatically carry over to the estimation of discrete choice models (on the SP data), see comments on exogenous sampling above.

## Exclusion criteria, Experimental designs and Estimation

In Section 5.1, Ian Williams discusses the implementation of exclusion criteria. In particular, he highlights that criteria 2-4 have not been implemented. As a result, a substantive number of responses is retained from respondents with reservations regarding the realism and/or their understanding of the context or questions that were used in the stated preference experiments.

Our decision to only implement exclusion criteria 1 (removing non-rational respondents) and 5 (respondents accepting a very high boundary value of travel time) but not criteria 2-4 was carefully documented and discussed with the client before proceeding to the final model specifications. In general, it is not feasible to estimate all models across all possible exclusion criteria and hence we decided on the most appropriate ones early on, and continued model development from there.

Indeed, we would have liked to use a sample of respondents with a full understanding of the survey format and responding in a fully informed manner. Unfortunately, a balance had to be struck between retaining a large enough sample for estimation, and high-quality responses. Overall, the influence of the non-implemented criteria 2-4 was limited in the early stages of analysis – as discussed on page 19 and Appendix A.3 of the final report - and our intuition is that this is unlikely to change the overall recommendations. This is particularly true for the OGV samples, but perhaps less so for LGVs. Even in the latter case exclusion criteria 5 was found to be most influential.

Section 5.2 of the review reflects on the presented levels for transport costs, and the implicit boundary values of travel (BVTT) time put in front of the respondents.

In each choice task, the respondent is asked to make a choice between two route alternatives A and B, each described in terms of a travel time (T) and cost (C). The ratio of the cost and time difference between these two alternatives is called the boundary value of time (BVTT):

$$BVTT = -\frac{C_B - C_A}{T_B - T_A}$$

The BVTT is expressed in £ pounds per hour. Presenting respondents with a range of BVTT values aims to tease out respondents' willingness to pay to reduce travel time, and thereby the presented values do not necessarily coincide with actual cost per hour. Namely, if we expect a low VTT (e.g. for shippers that contract out), we present a range of low BVTT values to narrow down the respondent's VTT as accurately as possible. Presenting these respondents with very high BVTT responses would result in only (or most) choices for the slow but cheap route which provide limited information on the actual VTT underlying these decisions.

The review correctly highlights that across the different segments (e.g. OGV1, OGV2, LGV; shipper, carrier) the presented BVTT values are very distinct. This has been implemented by design and informed by the international literature. For example, low BVTT bids for shippers were chosen because in the international literature typically 85% of total VTT was attributed to the carrier/time-dependent transport costs and 15% to the cargo. The transport cost component of the VTT, i.e. the values elicited from carriers, in several of these studies was found to correspond well with the full transport cost per hour minus the pure distance-related costs (that is mainly the fuel costs). Similar (low) BVTTs for shippers that contract out were implemented in the latest national freight VTT study in Norway, and in the referred study these expectations were confirmed (Halse et al. 2019).

Specifically, in Section 5.2.8 the choice for working with BVTT ranges specified in £/ton/hour for carriers is questioned and an alternative unit of £/vehicle/hour is suggested. We had not thought of this, but it seems a good suggestion. On the other hand, our use of costs per tonne is not unusual. Prices for shippers are often stated per tonne and many transport models also use cost functions with cost per tonne (e.g. for mode choice). The idea here is to introduce economies of scale in transporting goods: unit costs per tonne are lower for bigger vehicles than smaller vehicles.

In relation to the BVTT measures presented to shippers (see Section 5.2.9 of the review) some additional clarifications are made below which should have been included in the main study's final report:

- In the SP experiments, the link between the different types of carriers and shippers and the lines for the BVTT range in FR Table 45 was made as follows:

    - If Q0=Haulier-> use first line (carrier) from Table 45 for BVTTfac.
    - If Q0=Shipper(a) who outsources ALL their transport -> use line 3 or 4 from Table 45 (shipper hire-and-reward, LGV or HGV) for BVTTfac.
    - If Q0=Shipper(b) who PARTLY outsources transport AND B0=Own account -> Use line 2 from Table 45 (shipper own account) for BVTTfac.
    - If Q0=Shipper(b) who PARTLY outsources transport AND B0=Subcontract -> Use line 3 or 4 from Table 45 (shipper hire-and-reward, LGV or HGV) for BVTTfac.
    - If Q0=Shipper (c) who does NOT outsource ANY transport -> Use line 2 from Table 45 (shipper own account) for BVTTfac.
- Some changes were made directly in the programme after the questionnaire (in Word) had been developed and the Word questionnaire (that appears in the report as A.2) was not updated to reflect these changes. We apologise for this omission. The actual programme works as follows:
    - In Stated Preference setting 1 (SP1), which covers time vs. cost choice tasks, we give the instruction:

"*If in some alternative the journey time would be shorter than it currently is, then you could possibly use this time gain to employ the vehicle and driver elsewhere in a productive way (and with longer times you possibly have additional costs for the vehicle and personnel).*"

    - This instruction is given to:
      Carriers/hauliers (Q0=1)
      Shippers C who do not outsource any of their freight transport (Q0=4)
      Shippers B who partly outsource their transport (Q0=3) AND who have a selected transport for which they carry out the transport themselves (B0=1, own account).
    - The other shippers B have a selected transport that they contract out (B0=2, subcontract). This group did NOT get the above instruction to think about the driver and vehicle time. The same goes for shippers A who outsource all or their freight transport (Q0=2).

Overall, the instructions shown depended on both Q0 and B0, not only on Q0 (as was incorrectly written in the Annex of the final report). This also applies to Stated Preference setting 2 (SP2), which covers choice tasks involving trade-offs across time, cost and reliability. This should address the concern that "*much of the differentiation appears to have been erroneously generated*" (Section 5.2.10).

Zooming out from the detailed discussion on the experimental design, the review provided a welcome opportunity for us to check the implementation of the different cost components to ensure the experimental design matched our original intentions. Besides acknowledging that our explanations could have been clearer and more complete in the

final report, we have not found errors in allocating the right BVTT range to the respective segments.

This leaves us in a situation where we do not have a clear explanation for the patterns in the share of acceptance of the BVTT. We agree that it probably is related to us presenting lower BVTTs for shippers that contract out, but this design was in accordance with the findings in most of the international literature. The information that we now get for the UK from shippers that contract out (higher than expected VTT) suggests (in retrospect) that a design with similar BVTT values for shippers vs carriers might have worked better.

Section 6 of the review presents a range of helpful modelling suggestions and sensitivity tests which can be implemented in a later stage. Whilst recognising that the limitations of the final model specifications, specifically in terms of included explanatory variables, are largely informed by the limited sample sizes, we are pleased that the review has not identified significant limitations in the adopted approach to modelling.

## Implementation

Section 7.1 of the review expresses a concern regarding the double counting of the transport cost element, and thereby the risk of greatly overestimating the correct VTT and VTTR values. Ian Williams argues that this is largely driven by the respondents' ability to conceptually separate the product costs related and transport cost related aspects of travel time savings. Whilst respondents may have struggled with the interpretation, there are no clear indications in the open-ended responses to the study follow-up questions that his has been the case. Furthermore, we believe some of these concerns have only arisen at the end of the survey. The interview programme instructed the kind of costs that respondents should consider not only on being shipper type A, B, or C, but also based on outsourcing (or not) of the selected shipment. As discussed earlier in this response, the instructions relating to the outsourcing of shipments used in the programme were in fact different to those reported in the final report (and used as the basis for the peer review), and we therefore believe that the concerns expressed by Ian Williams are not as extensive as expressed in the referred paragraph. Table 8 in the review document is closely associated with this point and hence we do not fully agree with the judgement reached regarding the usability of the estimated values for shippers in TAG. This does, however, not take away the risk that some (but not all) respondents may have responded with a different definition of the relevant costs in mind than intended.

Sections 7.2 and 7.3 express valid concerns regarding estimating a representative overall average cargo value per journey tonne for an HGV type journey, especially since this average value across road links will be associated with a large variance. This discussion reverts to the preceding section on segmentation and representativeness, whilst it may be of interest to identify all this potential heterogeneity its implementation in practice is hampered by the difficulty to collect a dataset which is sufficiently rich (both in terms of

SP and revealed preference terms), the limited heterogeneity accounted for in existing transport models and the need for proportionality in appraisal practices. As such, the client required a mean value for implementation purposes.

In Section 7.4 several references are made to the non-representative nature of the collected sample. The reasons for not collecting a representative sample – and why that is non problematic - have been expressed in an earlier section. The latter comments in Table 8 are therefore not considered to be problematic in our opinion. Similarly, the adjustment suggested in 7.4.8a to conduct weighted regression is not considered desirable, especially because there is a risk of overweighting some specific individuals in the sample.

## Overall recommendation

Ian Williams ends Section 7.4 recommending not to replace the current TAG approach due to many caveats in the main study. As articulated in this document, we do not agree with all the caveats and support our values as the most likely VTTs. Specifically, we have highlighted that an informed approach to segmentation of freight transport, sampling, study design, and analysis has been implemented. The decisions made in each stage of the study have been informed by the broader national and international literature on freight values of travel time and discussions with the Department for Transport and National Highways.

Indeed, we would have preferred to work with larger sample sizes and have a better understanding of some of the emerging descriptive statistics and response patterns in the stated preference exercises. This does not take away that for road freight transport the main study is larger in size than other national VTT studies which have followed similar design approaches, and from which the results have been implemented in national guidance. It does articulate some of the challenges with surveying the freight sector in general. Based on a range of model specifications and testing of exclusion criteria, we believe that particularly for OGVs a set of robust set of estimates has been obtained. For LGVs this is perhaps less the case. As such, we recommend updating TAG and go beyond the currently included driver costs and Vehicle Operating Costs (VOC). As a minimum, TAG could include all the transport cost in the VTT and VOC. The latter has already been found for carriers in several countries. The estimation results for carriers in our final report largely confirm this for the UK. If one does not feel comfortable with the cargo time values presented in the final report, one might add a component for the cargo time costs to this based on the literature. This would, however, involve transferring values from other countries, such as France. In this regard, we wish to provide two reflections. First, internationally transferring values requires suitable mechanisms to correct for the differences across the economies (e.g. income) and the freight industries (e.g. types of goods, types of journeys, and vehicles) to avoid potential bias. Even when suitable transfer functions have been identified some degree of bias remains. Second, the carrier

OGV VTT value obtained is already quite high, and adding an additional cargo component on top forms an even larger departure from current TAG practices.

## References

Ben-Akiva and Lerman (1985) Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press.

ITS, Andrew Gordon Ltd, Intersection Analytics Ltd, Peak Economics Ltd (2025) Updating the Value of Travel Time: Scoping Study. Report for the Department for Transport

Halse, A.H., C.S. Mjøsund, M. Killi, S, Flügel, G.N. Jordbakke, I.B. Hovi, M. Kouwenhoven and G.C. de Jong (2019) The Norwegian value of time study on freight transport 2018 (in Norwegian), Transport Economics Institute, Oslo.