

# Sampling proposals for the Travel and Tourism Reform Project: a note for MARP, July 2023

Gareth James and Owen Abbott

## 1. The request from MARP

This brief note follows discussion at the MARP meeting of 20 June 2023 about the paper by Shenhuy and Davies ([Confluence link](#)) on plans for Travel and Tourism statistics, which can be referenced if further details are required. In this paper, we ask for further feedback from MARP on the proposed data collection and sampling strategies, which are necessary to meet budget constraints.

To assist in this request, in Section 2 we briefly describe the proposals again. Then, in Section 3, we have expanded and focussed the discussion and have inserted related, example questions into the narrative to help guide feedback on particular areas. Additional feedback would help provide us with assurance that we are proceeding in a reasonable and justifiable direction with our proposals.

In the meantime, and planned for the coming months, is further research and development of the proposals. We will return to MARP again with additional details and results when available.

**In summary, in this paper we ask MARP for assurance on the direction proposed for sampling and data collection for travel and tourism statistics and for any other thoughts on the potential risks in the approach. More specific questions for consideration follow, and are presented in boxes in the text.**

## 2. Background and summary of the proposals

This work is part of ONS's Travel and Tourism Reform Project (TTRP), which aims to improve the quality of travel and tourism (T&T) statistics as well as realising some necessary cost savings for the office, as the International Passenger Survey (IPS) in its current form is no longer affordable.

In this paper we focus on the collection of data about international travel and tourism – principally destinations, lengths of stay, and expenditure – and its collection via sample surveys. (We note that this work is not about migration statistics.)

Currently, ONS's International Passenger Survey provides data for T&T statistics, with interviews taking place with both departing passengers and arriving passengers at various UK air- and seaports (and also on particular sailings) and with respect to Channel Tunnel operations. For brevity, we will concentrate here on proposals for just the air-travel sampling, with those for other modes of transport being similar in nature or with less difference from current practice.

Two notable changes are envisaged under TTRP with regard to ONS's surveys:

**First** is a new Departures survey, which will replace the IPS Departures survey. Its main purpose is to collect:

- spend in the UK by overseas residents plus details of the trip undertaken.
- pre-trip spend by UK residents plus details of the overseas trip just beginning

To survey most effectively, we are proposing to continue sampling departing passengers at their point of exit from the UK (i.e. at airports, as per the IPS), but with some changes to the sample design and the location of data collection within the airport. In particular, instead of selecting passengers for interview via a systematic sample as they exit the security checks area, we will select flights as sampling units (clusters) and then select passengers on those flights for interview at the boarding gate. The overall proposed sample size of the new survey will increase (size to be confirmed, but up to 25% larger than that of the current IPS Departures Survey has been suggested), which is possible from the savings made by not running an Arrivals survey (see Second change, below).

In addition, the Civil Aviation Authority (CAA) also conducts a survey of departing UK and overseas passengers at (some) UK international airports. The CAA survey started in 1969 and more recently has incorporated a core set of questions in common with ONS's. Our proposed, new ONS Departures survey also has a sample design that is congruous with that of CAA's survey, which means we can pool the datasets; the resulting sample will be notably larger than the current IPS, which should result in much greater precision in estimates.

In addition, the CAA survey doesn't appear to suffer from the issues of "imbalance" (differences between estimates of arriving and departing passengers for a given destination) present in IPS data. We believe this issue is likely mitigated by sampling flights as clusters and interviewing at boarding gates, giving an additional advantage to this proposed change of approach.

**Second**, we will stop ONS's IPS Arrivals survey, which is a necessity to achieve the cost savings ONS requires. In its place, the required travel-and-tourism-related questions are being added to VisitEngland, VisitScotland and Visit Wales' combined Great Britain Tourism Survey (GBTS) and Great Britain Day Visits Survey (GBDVS) to collect data from UK nationals about their recent trips abroad. Modelling of that expenditure of UK nationals on overseas trips will follow, and those models can then be used to predict the spend of UK-resident responders to the combined ONS + CAA departures survey data.

Thus, the combination of modelled/predicted expenditure with the Departures data on destinations and lengths of stay will allow total expenditure of UK residents to be estimated across the various output domains required.

### **3. Discussion and additional questions for MARP**

We now discuss four principal differences that will result from the proposed approach for travel and tourism statistics, and on which we would appreciate further feedback from MARP.

At the end of each section, we have added example questions on the topic for MARP to consider but would welcome any additional thoughts around these topics.

## **(a) The selection of flights as Sampling Units (Clusters) for Departures**

Airport terminals are essentially the main form of stratification in the sampling scheme, with a selection of time intervals (these are c.8-hour shifts that pairs of interviewers will work) then sampled each quarter or month as the primary sampling unit (PSU). Specific flights within those shifts will then be selected as secondary (or possibly later-stage) sampling units. Within each selected flight, a sample of individual passengers will then be selected as the final sampling units for interview at the departure gate.

Use of flights as secondary sampling units, which are essentially clusters of passengers, will likely be less statistically efficient than drawing a simple random sample, which is how the IPS currently operates via systematic sampling of all passengers on any flights after they pass through security. A notable design effect could be introduced here as, intuitively, passengers on the same flight will be more similar to each other with respect to target survey variables than to all passengers more generally. However, the increases in sample size (on this survey alone and when pooled with CAA data) will help mitigate those effects. In addition, the current approach is known to suffer from differential non-response concerning destination/origin of passengers, and thus sampling of flights with interviewing at the departure gates may help reduce the potential for bias from this.

Further work is required to investigate the effect of this additional stage of sampling by understanding the within-cluster and between-cluster variances. Considerations will also include options around the number of flights selected vs number of passengers per flight (or the sampling interval, whichever is specified), though practical considerations will heavily constrain what is possible. This work may not be particularly easy to undertake either, as historical data naturally come from the particular sample design used at that time, but we do have IPS and CAA data collected under different designs that can be compared, and simulation studies may be possible, resources/time permitting.

We may find that additional stages of sampling or stratification (destination region, routes, gate or groups of nearby gates) may be required to support the balance between practicality and precision. Implicit stratification through systematic selection, and selection Probability-Proportional-to-Size will also be considered when selecting and allocating samples across airports, time and shifts. The effect of such design-decisions, combined with practical factors such as interviewer availability and ability to move between gates, etc. when required, means there could be wide variation in the final sampling weights, which will require careful investigation.

Our immediate tasks are to develop and compare designs and to specify the details of the proposed ONS Departures Survey. We can return to MARP with that additional detail in due course.

### **Questions for MARP**

Does MARP have any comments on the expected advantages and disadvantages of this approach to sampling (further detail was given in the previous paper)?

Are there particular risks or other aspects we should consider further in our investigations?

## **(b) Practicalities of sampling and interviewing passengers at the boarding gate**

Whereas implementing a 1-in- $k$  systematic sample of passengers as they exit Security at an airport is relatively straightforward (in theory, anyway), there are different challenges in sampling passengers gathered at boarding gates.

At best, systematic samples could be drawn if passengers are queued and stationary, but it seems likely that various combinations of priority queues, non-priority queues, seated/standing passengers in flight-specific areas and in more communal areas (i.e. mixed flights) together with a mix of early- and late-arrivers will all be encountered by interviewers at different times and in settings. It seems a best-efforts approach will be required along with new interviewer-guidance on how best to draw a random (or quasi-random) sample. A recent ONS trial (internal paper) of sampling at boarding gates, along with this being CAA's current approach, has however shown that the approach is workable.

Those recent trials have proved positive in terms of data collection too. The observed non-response/refusal rate was lower than that of the IPS, with good survey engagement observed. The intuitive explanation lies in passengers likely wishing to be elsewhere after exiting Security – heading straight to the boarding gate, for pre-departure food/drinks or for some airport shopping probably all appeal more than stopping to answer IPS survey questions! By contrast, (most) passengers arrive at the boarding gate having completed all other airport-processes and an ONS survey may even provide a welcome distraction while waiting to board the aircraft. One potential issue noted by interviewers in the trial though lay in concerns about privacy amongst fellow passengers: spend being overheard, etc. was cited in a few cases. This could be mitigated with on-tablet visuals.

Overall, we believe interviewing at the gate is achievable in a sufficiently random way (with new interviewer-guidance to be produced) and will bring benefits in terms of response- and engagement-rates with the survey and help with the “imbalance” issued noted previously. Nonetheless, we will need to accept some trade-off here determined by what is practicably achievable.

### **Questions for MARP**

To what extent should we be worried about deviations from theoretical practice on that final stage of sampling?

How much impact are deviations from ideal sampling practices at the gate likely to make, and to what extent might those be mitigated by achieving better response rates (with potential for less non-response bias)?

Are there particular aspects of that final stage of sampling that we should insist are carried out and can include in our guidance for interviewers?

### **(c) Pooling ONS Departures Survey data with CAA Departures survey data**

Although its design should be analogous to ONS's proposed design, we do not have control over the CAA survey design nor its operations. We may be able to influence aspects of the design and operation through discussions with CAA, however, and will have a service-level agreement in place. Nonetheless, we will have to accept the risk of alterations to sample processes or deviations from expected practice and any consequences these might have on survey error (bias and variance).

The main benefit of pooling the datasets is the notable increase in sample size, especially for the core (common) variables, which should see a commensurate increase in the precision of resulting estimates. Beyond that, estimation of non-core variables (those not in common between the surveys) should also benefit from the larger overall sample size through use of composite calibration methods, such as those in Merkouris *et al* (2023). The potential gains will be explored through further research.

#### **Question for MARP**

Please would MARP advise on the benefits, risks and best approaches (including methods) on how to pool datasets in the way proposed?

### **(d) Modelling expenditure overseas using household survey data**

Ceasing the surveying of Arrivals at UK (air)ports provides the cost-saving that ONS requires but presents the challenge of how else to collect that data, in particular the overseas-expenditure information from UK residents. Our proposal is to use a household survey – namely VisitEngland, VisitScotland, Visit Wales' combined Great Britain Tourism Survey (GBTS) and Great Britain Day Visits Survey (GBDVS) – for this purpose, asking respondents details about their most recent overseas trip in the past few weeks (with the length of 'few weeks' to be defined).

Some points to note for this proposal are that the GBTS and GBDVS are currently online panel surveys, rather than probability samples, although quotas and weighting provide representativity according to a number of metrics. Between them, those surveys currently have an annual responding sample size of 95,000 across England, Scotland and Wales. Not all of those respondents will have undertaken a qualifying trip, though we expect about 25,000 to have done so.

Our intention is to fit a model to the GBTS and GBDVS data for the purposes of predicting overseas spend of UK residents. The exact model type and approach are yet to be determined but could use multiple years to increase effective sample sizes and potentially provide feedback on important modelling variables to collect in the ONS and CAA surveys. If possible, the models will then be applied to the departures survey data to predict expenditure abroad for the respondents. This would result in a dataset which be the base for all Travel and Tourism outputs. Alternatively, if the models do not work well for prediction, we would explore aggregate-level estimates. We are planning to use existing IPS data to explore this approach (for instance by simulating a GBTS or GBDVS type sample).

We are also currently investigating the accuracy of respondents' recall of expenditure data on their most recent trip, as well as testing, assessing and

developing models and the overall approach. Our hypothesis is that recall may be better in the household panel survey than a port-based survey. In future work we will also be considering whether administrative data can play a role here, either directly or as a quality check. We can return to MARP with further information as our investigations develop.

### **Questions for MARP**

What are your thoughts on the risks of this approach? Given the need to make survey cost savings, does use of an existing household survey – rather than an Arrivals surveys conducted at a port – together with modelling of expenditure seem something we should work towards?

Does it matter that the proposed household survey is (currently) a non-probability panel survey, given it will be used only to derive models? Does it matter that it is online only?

What concerns do you have about deriving predicted overseas spend via models fitted to these data?

Do you have a preferred approach, or suggestions, on how best to use the modelled data and departures data in conjunction?

Do you have suggestions for how we should ensure and maintain quality in the future if this approach is taken forward, as we are planning?

### **References**

Merkouris, T., Smith P .A. and Fallows, A. (2023) Combining National Surveys with Composite Calibration to Improve the Precision of Estimates from the United Kingdom's Living Costs and Food Survey, *Journal of Survey Statistics and Methodology*, Volume 11, Issue 3, June 2023, Pages 713–741, <https://doi.org/10.1093/jssam/smad001>

Shenhuy, B. and Davies, K. (2023). Travel and Tourism Reform Paper. Methodological Assurance Review Panel. Discussed at meeting at 20 June 2023. <https://confluence.ons.gov.uk/download/attachments/36206105/EAP199%20-%20Travel%20and%20Tourism%20Reform%20Report.docx?api=v2>