

DUPLICATION OF READING IN PRINT AND ON-SCREEN

How is duplication between reading in print and on screen controlled?

Duplication of reading is largely a product of the fusion process, with a final calibration to targets derived from data collected from the PAMCo Panel.

What is the PAMCo Panel?

The PAMCo Panel was set-up to provide single-source data on the proportion of readers reading publisher brands in both print and digital. This was a major innovation in the PAMCo set-up, as such data had not previously been available at scale, and there was no reliable benchmark for duplication levels.

Around 5,000 survey participants per annum agreed to have a tracker app installed on all their digital devices to track their reading of publisher brands for 28 days, following on as soon as possible after their PAMCo readership interview. Panel recruitment commenced in April 2016 and ran until December 2018, during which time almost 10,000 panellists had their digital reading behaviour tracked.

The sole purpose of the PAMCo Panel was to provide the best possible benchmark for print and digital duplication of reading. It was never used to derive digital audience estimates, which come from a data fusion with Comscore. Indeed, with an annual sample of 5,000, the Panel would have been too small to provide direct measures for more than a small number of brands

Why did Panel data collection not continue?

The Panel data confirmed that the fusion Ipsos carry out to integrate Comscore's digital audience estimates with print readership estimates from the PAMCo survey is highly successful in capturing duplication of reading between print and digital.

In most cases adjustments to match the 'duplication targets' generated by the PAMCo Panel result in monthly brand reach which is within $-/+2\%$ of that which would be estimated by fusion alone. As such, the Panel validated the fusion process and it was decided by the PAMCo Board and Technical Group to cease Panel data collection with effect from January 2019.

Adjustments to print and digital duplication levels by brand continue to be made based on the Panel data, and in time may be supplemented by a modelling procedure to take account of trends if necessary. There is no discernible change in the PAMCo data as a result of ceasing Panel data collection.

How are the Panel data used?

The Panel data indicate how much more likely print readers are than non-print readers to visit digital platforms of the brand concerned. For some brands there is very little difference between print readers and non-print readers, for others there is a much stronger relationship, with print readers being much more likely than non-print readers to visit digital platforms for that brand.

This information is used to set a target print and digital duplication for each brand, which must take into account both the information from the Panel and the size of the Comscore digital audience estimate. Duplication targets are set at a monthly level, and initially for digital reading as a whole (though the allocation process takes into account the various different digital platforms).

The target duplications derived from the Panel data are used to adjust the print and digital reading duplications in the fused dataset and ensure duplication is represented fully. For the largest brands these adjustments tend to be minimal, as the fusion is already capturing most of the expected duplication. For

smaller brands around 20-30% extra duplication (relative to the fusion alone) is added in, but even in these cases the effect on total monthly brand reach is small.

Was a sample size of 5,000 enough to generate duplication targets for the smaller brands?

Ipsos use a sample of around 5,000 to generate the duplication targets, though in practice samples by individual brand may be less than that if tagging has not been at 100% throughout. Ipsos developed an innovative solution to provide duplication targets for all brands:

- A smoothing approach is used which is variable by sample size. For large sample brands smoothing is minimal, while small sample brands have more smoothing applied to their targets, to avoid volatility.
- The smoothing model also draws on claimed duplication data from the PAMCo print survey to help predict duplication targets for brands for which there is insufficient panel sample (though it would not be appropriate to take these data at face-value due to the biases and understatement in what participants claim they have read on screen).

Why are the duplication targets calculated, rather than taken directly from the Panel?

- To take account of different levels of readership in the fused PAMCo/Comscore database compared the PAMCo Panel, e.g. if the Comscore estimate is bigger than the digital reach observed in the Panel, which is often the case.
- To take account of sample sizes which are too small to produce reliable duplication estimates.

How representative was the Panel?

The Panel was recruited directly from the high-quality PAMCo sample, with every effort made to encourage participation. The requirement was to measure visits to publisher sites across all the participant's devices, rather than a single device, which would be much easier to achieve. As such, the Panel quality is high relative to the more usual 'opt-in' sample used for digital measurement panels.

As with most samples, there are some skews in the unweighted sample. The most noticeable is to somewhat under-represent those with 3+ devices, and a tendency to under-represent men and over-represent women. The Panel data are therefore weighted by gender and age, newspaper and magazine readership (to ensure print readers are not over-represented) and 4-week print readership and device usage, to correct these skews as far as possible.

Other points to note are:

- Reading via work computers is not included. It is not possible to achieve the necessary permissions to provide a representative sample of reading at work for PAMCo's purposes.
- Reading via third-party platforms is not included, other than for Google AMP if the correct tags are in place.
- The Tracker App cannot identify if different people are using a single device, particularly as it is designed to be as non-intrusive as possible. Only primary users of phones and tablets will have their reading measured, which helps reduce the likelihood of shared usage. While in some cases there will be some overstatement of digital repertoire, the claimed usage data suggest relatively low levels of shared usage.

The PAMCo Technical Group conducted an assessment of the data provided by the Panel, which are credible in terms of relationships between the print and digital platforms, and between different types of brand and content.

Are PAMCo likely to need to set-up another Panel to measure duplication of reading between print and digital if digital reading behaviour changes over time?

This would only be necessary if there were very significant changes, for the following reasons:

- A key factor determining the proportion of duplication in reading between print and digital is the relative size of the platforms, which are tracked by the ongoing audience measurement of PAMCo and Comscore. For instance, if audiences reading via pcs drop this will be measured by Comscore, and when these new Comscore pc audience data are incorporated within PAMCo the fusion will reflect a lower proportion of print readers reading via pc.
- It is important to remember that the duplication data generated by the Panel are based on monthly digital reading as a whole, and for smaller brands the sample sizes available mean duplication targets must be modelled, drawing in part on survey data. Even if the Panel had continued there is a limit to the sensitivity with which it is possible to track trends for small brands, individual platforms, particular demographics etc
- It is possible that there could be more structural changes in duplication over time, though this requires major and sustained change. If, for example, the digital audience for a brand experiences a large and sustained increase, it may be that the 'new' digital readers are less likely to read in print than the longstanding digital readers. The possibility of structural change will be checked by monitoring the overlaps in the claimed survey data. Claimed data are not a good guide to absolute levels of duplication, but will give an indication if there are significant trends which need to be taken into account. Ipsos will also be monitoring the amount of calibration in the fused dataset, as an observed increase may indicate underlying change.

If there are indications of fundamental underlying change the plan is to extend the use of claimed survey data in the duplication modelling process, where it already plays a role.