



Data Fusion In a Nutshell

Bite Sized White Paper | **2011**

What do the following have in common:

Rock and roll, the Sun, a slipped disc and chicken tikka masala?

The answer is of course 'fusion', an adaptable word that conveys quite different meanings to a musician, a physicist, an orthopaedic surgeon and the man on the Boris bike. Our usage, data fusion, refers to merging two or more surveys by matching up respondents on common variables, or 'hooks', such as demographics, lifestyle and media and product preferences. It has proven most useful for combining media exposure with product consumption data, providing valuable insights for media planning.

The main attraction of fusion is to avoid the commercial and technical difficulties inherent in capturing all the data of interest in a single-source survey.

How we got here

Fusion has a long and chequered history. Like so many techniques it started in the military, combining data from arrays of radar dishes. Then the medical profession developed it for their data clusters, such as ECG readings.

One of the earliest adopters in media research was Friedrich Wendt at AG.MA (Association





for Media Analysis) in Germany, who was developing applications in the seventies, based on linear programming like the military.

Wendt used many terms that have become essential: single-source, donors and recipients, common and specific variables, distance, marrying individuals.

The '80s saw an influential French school, led by Gilles Santini, which developed several new algorithms.

The first major British contribution came in the early '90s. The Market Research Development Fund backed a test in which one half of the TGI was fused onto the other – now known as folding. This is a powerful form of validation which allows the predicted selectivity of media against target audiences to be compared to the actual values. This work established a new and relatively simple algorithm which has stood the test of time:

Find the best suited couple and marry them, then the next best, and so on. If there are too many maidens or bachelors, allow limited polygamy or polyandry...

After one or two false starts, the first successful commercial application in the UK was the fusion of TGI with BARB to form the Target Group Ratings. TGI was fused onto BARB rather than the other way round, so as to preserve the ratings currency. Though this was a purely commercial decision, recent methodological advances have shown it had technical merit too.

After a long life, TGR has now been supplemented by IPA TouchPoints, a consumer-centric, multi-media database launched in 2006, with new releases in 2008 and 2010. A hub survey of 6,000 adults is enhanced by fusing on several major media surveys, including BARB (panel and Establishment Survey), NRS, RAJAR and soon POSTAR.

On the other hand

The success of these applications – and others overseas – came in the face of **the many challenges presented by fusion**, both technical and commercial:

- There is no dominant methodology, like say k-means for cluster analysis. Indeed, according to Roland Soong (formerly with KMR), there can be no single best method.
- There is no out-of-the-box algorithm for any methodology. Much tender loving care is needed in preparing the datasets, choosing the hooks etc.
- Fusion only preserves that part of the relationship between the specific donor variables and the specific recipient variables that is accounted for by the hooks.
- By definition, a fusion requires purveyors in different fields to come together and agree on the project, its methodology and its funding.



- Some of the methodologies are difficult to describe for the layman, so appear to be black boxes.
- Fusions are not automatically successful. This leads to risk for sponsors and the potential need for independent assessment.
- Rival methodologies that claim to offer the same or greater benefits with less complication and cost have emerged.

These issues underpin fairly widespread residual concern over the validity of fusion, which has limited the determination required to develop a fusion and bring it to market in many countries.

The future is ...

We think these concerns have been overstated. **The methodology has come a long way since the early days.** The metrics of validity are now well-known, particularly regression to the mean: the dilution of selectivity indices due to poor hooks or matching. The means to quantify them are in place, such as folding and limited single-source comparative data. A long-running debate as to whether it is better to fuse 'small samples onto large' or 'large onto small' was recently resolved by results in a paper in IJMR by Trevor Sharot, Consultant Statistician to Ipsos MORI, that allow the margins of error for any potential design to be





calculated in advance. Among other results, this paper established that it is always better to fuse the larger survey onto the smaller one.

In short, while constructing a fusion remains hard work, it is no more fraught than any other form of data-modelling. In these times of media convergence, with a wide and growing range of devices and channels being offered to the public and a growing need to see the whole picture, fusion's time has surely come.

The full version of this paper with a list of the above references may be accessed at:
<http://www.ipsos-mori.com/fusion>

For more information, please contact:



Trevor Sharot

t: +44(0)20 8861 8217

e: trevor.sharot@ipsos.com

www.ipsos-mori.com

