

Use of Probabilities in PAMCo Bureaux Data

In the PAMCo data, whether a person has read or looked at a publication in print or online is expressed in one of two different ways, either as a binary (1/0) variable or as a probability. Traditionally, probabilities have been used as the required input for reach and frequency algorithms and binary data has been used within the cross-tab systems but in the PAMCo dataset (as in its predecessor NRS PADD) probabilities have also been used for some parts of the data within the cross-tab systems because daily and weekly reach in comScore's datafile take the form of probabilities.

Cross-tab software

The following table summarises where reach is held as binary data at the respondent level and where it is expressed as probabilities. An explanation follows the table.

Platform	Daily Reach	Weekly Reach	Monthly Reach
Print	Binary	Binary	Binary
Digital	Probabilities	Probabilities	Binary
TBR	Probabilities	Probabilities	Binary

In all cases (Print, Digital and Total Brand Reach) monthly reach is expressed as a binary variable and can be cross-tabbed in the bureaux cross-tab software to produce entirely expected results.

Daily and Weekly Print Reach are also held as binary variables and cause no unexpected results in the bureaux systems when looking purely at Print audiences.

Daily and Weekly Digital Reach and Total Brand Reach are expressed at the respondent level as probabilities. At a simple level, when looking at a single item (including a single TBR) on a single platform, results in the cross-tab systems will produce expected results.

However, users should be aware that when they combine digital entities or TBRs or platforms when cross-tabbing daily or weekly reach, they will see unexpected results that are not "correct".

For example, trying to replicate Total Brand Reach definitions by specifying the component parts (using “OR”) will only work for monthly reach but not for Daily or Weekly Reach, as the overcode for TBR has been calculated at Ipsos using the appropriate mathematical formula, which is not replicated in the bureaux software.

In the same way, users will not be able to replicate daily or weekly reach for the “Mobile” platform by specifying as Daily Reach for Phone OR Daily Reach for Tablet.

The table below illustrates that it is also inadvisable to cross-tab TBR with its component parts as this will produce anomalous findings in the weighted results (but not the unweighted results).

In this example, the weekly TBR for Brand X is 17,811,000, of which 12,935,000 are digital readers and 6,364 are print readers. But if users specify TBR as a column they will not see the correct weighted TBR or digital audiences in that column. The print audience will be correct because it is created from a binary variable and all unweighted audiences will be correct

		Total	Brand X (TBR) (D) - Total - Weekly Reach
Total	Audience(000)	52,818	17,811
	Resps	34,767	18,452
Brand X (TBR) (D) - Total - Weekly Reach	Audience(000)	17,811	14,079
	Resps	18,452	18,452
	%Col	33.7	79.0
Brand X (TBR) (D) - Total Digital - Weekly Reach	Audience(000)	12,935	9,202
	Resps	15,581	15,581
	%Col	24.5	51.7
Brand X (TBR) (D) - Print - Weekly Reach	Audience(000)	6,364	6,364
	Resps	4,764	4,764
	%Col	12.0	35.7

On the other hand, users can confidently combine platforms or entities with AND statements, knowing that they will see the expected audience. For example, specifying “**Brand X (TBR) - Total Digital - Weekly Reach**” AND “**Brand X (TBR) - Print - Weekly Reach**” will produce the expected overlap audience of 1,488,000.

Reach & Frequency Systems

It should be noted that in the Reach & Frequency Software the probabilities data match the standard (binary) data at the All Adults level.

However, if comparing probabilities data in Reach & Frequency programmes with standard Cross-tab data there may be small differences for audiences other than All Adults, with differences becoming larger as audiences become smaller.