THE PERSONALISED MEDIA LIST IN PRACTICE

Katherine Page, Ipsos Media

Synopsis

This paper reports on the results of a large-scale test of the Personalised Media List (PML®) methodology, a technique to reduce media list length by removing those titles likely to be least relevant to the respondent. The concept for this technique was first presented at the Symposium in Venice.

The results of the test indicate that application of the PML technique makes very little difference to the estimates obtained. This paper considers why that is so, and what the implications are for the way in which we handle media list length.

Introduction

In common with many other countries, there are concerns in the UK about the length of the media list measured by the National Readership Survey. Some users are concerned that the list of over 300 titles is too long, and that respondent fatigue may impact upon the quality of the estimates concerned. On the other hand, some users would welcome the opportunity to extend the list of titles surveyed, or the number of questions asked about each title.

There are a number of possible solutions to reduce the number of titles any particular respondent is asked about. One solution is to split the media list into a number of parts, and then show each respondent a selection of those parts rather than the entire list.

This, we believe, is the basis of the experiment ag.ma have been carrying out in Germany, which we will hear about at this Symposium. In the German test the media list has been split into three parts. Each respondent is then asked about two of the three parts. Great care is taken in deciding how to split the media list into these three parts, but the parts will be allocated to respondents at random. So, for instance, a respondent who reads motoring titles may or may not happen to be asked about the part of the media list that includes motoring titles. Another feature of the ag.ma test design is that the overall sample size has been increased by a third so that the base to analyse results for any particular title remains unchanged.

The PML technique, as developed by Ipsos Media, is a different solution. This technique personalises the media list for each respondent by attempting to predict which titles he or she is likely to have read. In essence, we filter which titles a respondent is asked about on the basis of a small number of demographic and topic interest questions at the beginning of the interview. The aim is to filter out the least relevant titles. In this way, the number of titles shown is reduced by around a third, though this varies by respondent.

We opted for the PML solution because:

It not only reduces the number of titles that each respondent is asked about, it also makes the media list more relevant to each particular respondent (and is unlikely to omit the titles they actually read)

The majority (90%+) of positive readership claims continue to be collected directly.

Given that it was unlikely that funding would be found to increase the overall sample size, it aims to make the best possible use of the existing sample (while accepting that, inevitably, if fewer respondents are asked about a particular title there will be an impact upon precision).

The filtering used for PML is complex. It is only possible with on-screen prompting, as it would be too difficult for interviewers to manipulate paper prompts. Since adopting Double Screen-CAPI (DS-CAPI) in 2003 the NRS has used on-screen prompting (whereby respondents are shown all prompt material on a tablet screen controlled by the interviewer's laptop). Once this technology was in place, it was decided to test the PML concept.

A split sample test of the PML methodology was conducted from April 2004 to June 2005. Alternate interviews were carried out with/without PML yielding a sample of 22,085 interviews carried out using PML and a control sample of 22,018 interviews carried out without PML.

PML worked exactly as intended, reducing the overall media list by a third. There was no evidence of any disruption to the data or of any biases introduced, the only negative being the expected impact upon precision. However, at the same time there is no evidence to suggest that reducing the Title Load per Respondent (or TLR, a phrase and acronym coined by Michael Brown) improved data quality. Some users hoped that the readership estimates might increase, and this has not materialised either.

This prompts the question of why we saw no discernible differences in the data when PML was tested? Before considering this question, we will give you an overview of the PML methodology and the main findings of the test.

The PML test methodology

From April 2004-June 2005 we carried out an on-Survey split sample test of PML, generating 22,085 interviews using PML and a control sample of 22,018 interviews without PML. The computer software selected every other interview as a PML interview from a random start (i.e. the sample was split within assignment and not by assignment). At the analysis stage, each half sample was separately weighted to ensure balance with regard to the key demographics.

Before outlining the methodology it is worth reminding readers that the British NRS uses the Extended Media List (EML) technique. (It was this technique which when introduced in 1984 enabled the media list to be expanded from around 100 to 300 titles, while at the same time substantially reducing the order effects observed, of which more later!)

The EML technique is as follows. The first time respondents are shown the titles, they are presented in groups of six on what are known as EML screens. The respondent is asked to identify those screens that show at least one publication they have read or looked at in the last 12 months. In this way the majority of EML screens will be eliminated from the interview before titleby-title readership questions begin. On average, each respondent selects ten EML screens (though of course this varies) out of the c.50 shown. The interview then proceeds by screening Read Past Year (RPY) readership for each title on each EML screen selected.

The PML methodology tested was as follows.

- 1. The 38 EML screens showing magazine titles were segmented into 26 clusters according to a small number of demographic and topic interest variables. The 'New Titles' screen/s and the nine EML screens showing newspapers were not included in the test. The variables used for the clustering were as follows:
 - Sex
 - Age
 - Topic interest
 - Future plans
 - Household Composition
 - Marital Status
 - ACORN (a geodemographic classification, taken from the sample information).

The 26 clusters identified were as follows:

1	Running
2	Young Music
3	Women's Style (3 EML screens)
4	Women's Weeklies (2 EML screens)
5	Health & Slimming
6	Hunting & Wildlife
7	Teen Girl/Pop Magazines (2 EML screens)
8	Upmarket Working/In-Flight
9	Fishing & Boating
10	Home Magazines (4 EML screens)
11	Supermarket Magazines
12	Store Magazines
13	Celebrity
14	Hair & Beauty
15	Golf & Photography
16	Bike Magazines
17	Football & Rugby
18	Car Magazines (4 EML screens)
19	Baby & Pregnancy
20	Wedding Magazines
21	General Reading
22	Gardening Magazines
23	TV Magazines (2 EML screens)
24	Men's Magazines (2 EML screens)
25	Science, Geography, History and Music

26 Empire, Viz, Time Out, The Big Issue

-

- For each cluster of magazines, past Survey data were analysed to identify the variables that were most powerful in 2. predicting whether a respondent was likely to have read any title within that cluster in the past year. Three strata were defined for each cluster of magazines. There was a different sampling fraction for each stratum.
 - 1 (Likely to read) Show all respondents the EML screens.
 - 2 (Unlikely to read) Show one in two respondents the EML screens. -.
 - 3 (Very unlikely to read)Show one in four respondents the EML screens.

This design meant that the majority i.e. c.90% of positive readership claims were still collected directly. The 'missing' 10% of readership claims (and the negative claims) from respondents in strata 2 and 3 who were not questioned were imputed from respondents in the same stratum who were questioned, as outlined at step 5 below.

- A small number of demographic and topic interest questions were asked before the readership interview commenced, 3. in order to ascertain the variables shown above. On the basis of these questions the computer software selected which stratum each respondent fell into for each of the 26 magazine clusters. This determined which magazine screens they were shown. If a respondent was selected to see a particular cluster of titles, he/she was shown all the titles in that cluster.
- Without further control, the number of magazine screens shown to each individual respondent would have varied from 4. 0 to 38. However, it was decided to show the TV magazine and running magazine clusters to all respondents, partly as a control and partly because there was no available back data on the latter. For the purposes of the test, it was also decided to restrict the range of screens shown so that no respondent was shown less than 15 magazine EML screens, i.e. the range was between 15 and 38 magazine screens.
- For each cluster, the sample collected in the lower two strata of respondents was used to impute readership claims 5. (largely in the negative) for respondents not asked about the titles concerned. At the client request, this was done so as to replicate the readership penetrations derived from those respondents who were asked directly about each title concerned. A special procedure was developed to do this, using a CHAID segmentation of the respondents who were asked about each title in strata 2 and 3.

Results of the PML test

The results can be summarised under four headings:

- Reduction in TLR
- Readership levels
- Order effects
- Impact upon precision

Reduction in Title Load per Respondent

TLR was reduced by a third. Instead of being shown all 38 magazine screens, each respondent in the PML half-sample was shown an average of 24 screens, plus the 9 newspaper EML screens, which were not subject to PML. (NB The exact number of EML screens varied a little across the duration of the test as titles left and joined the Survey).

Of course the degree of reduction in TLR was dictated by the specific design, i.e. how the strata were defined, the decision to ensure each respondent was shown a minimum of 15 magazine screens and so on.

The number of magazine screens shown to each respondent varied. The range was as follows:

Magazine EML Screens	% of respondents in PML sample				
15-20	29.1				
21-25	27.5				
26-30	26.3				
31+	17.1				

Because the selection of screens was based on the respondent's demographics and interests, it follows that certain sorts of respondents tended to see more (or less) EML screens. So, for instance, men saw on average three EML screens fewer than women, those aged 65+ tended to see fewer screens than younger respondents and so on.

One possible concern before we started the test was that this sort of variation in the number of titles shown to particular sorts of respondents might introduce a bias to the results. We found no evidence that this was the case, bearing in mind that there was very little evidence of difference between the PML and non-PML results at even an overall level.

As around 90% of readership claims are still collected directly with PML, it was not anticipated that there would be a substantial time saving in respect of the length of interview. This proved to be the case with a time saving of just 1.5 minutes arising from the fact that fewer EML screens are shown. (The PML interview took 25.1 minutes on average and the non-PML interview 26.6 minutes).

There was no strong feedback about PML from the interviewers. Where there were comments they tended to be positive, but not strong, and most felt that it made little difference, particularly because there was no marked decrease in interview length.

Readership Levels

The table on the next page summarises the readership comparisons for both Read Past Year (RPY) and Average Issue Readership (AIR). There are a number of features of these data to note.

The first is how similar the PML and non-PML data are, for both the magazines and newspapers. Looking first at the final column, gross AIR, we see that overall magazines subject to PML recorded 1% more AIR readership than in the control sample. Newspapers, which were not subject to PML, recorded 1% less readership in the PML half-sample.

Two magazine clusters (and the New Titles screen/s) were not subject to PML. These were the TV magazines and (for practical reasons) the running magazines. These clusters recorded 3% greater AIR readership in the PML half-sample, but these differences are not significant (using a two-tailed test and 95% confidence limits).

Indeed, none of the positive or negative differences for individual clusters at the gross AIR level were significant on the basis of a two-tailed test.

There is also no evidence that PML is screening in more occasional readers. One might hypothesise that these readers are more likely than regular readers to 'miss' reading claims (as supported by the evidence that order effects are usually more pronounced at the stage of screen selection and RPY than they are at the AIR level). If so, we might expect to see an increase at the level of screen selection and RPY claims, even if these extra claims from occasional readers do not translate into AIR claims.

However, when we look at the net RPY each screen of six titles (a close approximation of screen choice), there is no clear evidence of a difference. Furthermore, there is no evidence at all of an increase in gross RPY, as shown in the table on the next page. Gross RPY is 2% lower overall for magazines subject to PML, the same for the magazines not subject to PML and 1% lower for the newspapers, also not subject to PML.

So the readership levels recorded by the PML and non-PML data appear to be essentially the same at the both the RPY and AIR level both for the titles which were subject to PML and for those, such as the newspapers and TV magazines, which were not.

Summary of readership results from PML test

		Gross RPY	Gross AIR	
		Index	Index	
	Cluster	(PML/non- PML)	(PML/non- PML)	
2	Young Music	101	100	
3	Women's Style (3 EML screens)	95	97	
4	Women's Weeklies (2 EML screens)	100	103	
5	Health & Slimming	94	99	
6	Hunting & Wildlife	98	107	
7	Teen Girl/Pop Magazines (2 EML screens)	94	92	
8	Upmarket Working/In-Flight	100	106	
9	Fishing & Boating	107	130	
10	Home Magazines (4 EML screens)	98	101	
11	Supermarket Magazines	98	98	
12	Store Magazines ¹	96	98	
13	Celebrity	98	104	
14	Hair & Beauty	91	88	
15	Golf & Photography	96	99	
16	Bike Magazines	99	103	
17	Football & Rugby	104	106	
18	Car Magazines (4 EML screens)	98	103	
19	Baby & Pregnancy	106	114	
20	Wedding Magazines	100	100	
21	General Reading	101	104	
22	Gardening Magazines	103	98	
24	Men's Magazines (2 EML screens)	101	102	
25	Science, Geography, History and Music	99	102	
26	Empire, Viz, Time Out, The Big Issue	98	103	
	Total magazines subject to PML	98	101	
1	Running magazines	111	135	
23	TV Magazines (2 EML screens)	102	103	
	New Titles	96	102	
	Total magazines not subject to PML	100	103	
	Quality daily newspapers	99	98	
	Mid market daily newspapers	99	97	
	Popular daily newspapers	99	99	
	Quality Sunday newspapers	100	100	
	Mid market Sunday newspapers	98	99	
	Popular Sunday newspapers	100	100	
	Total newspapers (not subject to PML)	99	99	

 $[\]overline{1}$ Store magazines were only subject to PML for the first nine months of the test.

Order Effects

An examination of the order effects with and without PML is a key indicator of whether there has been any improvement in the quality of the data obtained.

Most of the concerns about excessive media list length hinge on the idea that presenting respondent with a large number of titles causes fatigue and this in turn depresses readership estimates as respondents either miss titles they have read or choose not to acknowledge them.

Evidence for this fatigue is usually taken to be the order effects, i.e. the relative difference in the estimate obtained for a title when it is shown early in the media list as opposed to when it is shown late. (Though, as Michael Brown has pointed out, "order effects' are not in fact synonymous with 'media list effects", as there are order effects that have nothing at all to do with questionnaire length.)

Nevertheless, the assumption is that the greater these differences, the more evidence there is of respondent fatigue.

Order effects are generally most apparent at the stage of screen choice and RPY, and less so when it comes to AIR. A hypothesis to explain this is that there is more likelihood that occasional readers (or at least those without a strong relationship with the title) will fail to record reading events.

Shown below is a comparison of the indices of the forward/reverse rotations on of the media list for the PML and Control samples for both RPY and AIR, based on unweighted data where an index of 100 is taken as the readership level recorded regardless of rotation. Data for screens 4-6 are omitted as these are allocated to regional titles, which are not consistent across the media list.

Publication (EML screen numbers)	RPY PML (Forward/Reverse)		Control (Forward/Reverse)		AIR PML (Forward/Reverse)		Control (Forward/Reverse)	
National dailies (1-3)	111	89	111	89	103	97	104	96
Magazines (7-16)	102	98	103	97	105	95	108	92
Magazines (17-24)	95	105	97	103	102	98	102	98
Magazines (25-34)	98	102	99	101	101	99	101	99
Magazines (35-44)	96	104	95	105	97	103	96	104
National Sundays (45-47)	91	109	91	109	97	103	98	102

Here again we can see that there is very little difference between the order effects observed within the PML sample and the non-PML sample, at either the RPY or AIR level. As expected, the order effects are greater at the RPY level, but this is equally true for both the PML and non-PML samples.

The data above are a somewhat bland in that they summarise the order effects across a number of EML screens. Of course different titles, and groups of titles, will be more or less susceptible to order effects depending on the profile of their readers, the saliency of their brand and so on. There are patterns hidden behind these summaries, for example the existence of order effects within a long run of titles on the same subject matter, such as the home interest titles. However, again there is no evidence that the PML sample is any worse or better than the non-PML half sample in this respect.

Impact upon precision

Because PML reduces the number of respondents asked directly about each title, it follows that the confidence intervals for the estimates obtained must increase. Calculations on the basis of the first six months of the split-sample test indicate an average increase in the confidence limits of around 30%. There will also be a further small impact as a result of the imputation process, which has not as yet been calculated.

Bear in mind, however, that this comparison is made against the existing methodology of asking all respondents about all titles rather than some alternative means of reducing TLR, which, if it involves splitting the media list between sub-samples, must also increase confidence levels unless a larger overall sample size is commissioned.

Why has PML had no apparent effect upon the readership estimates?

Given that, as we have seen, there appears to be no differences when using PML, it is worth reviewing whether this is in fact a surprising finding. We have three hypotheses as to why, despite being entirely successful in reducing the Title Load per Respondent, PML has not made a discernible difference to the estimates obtained.

Hypothesis 1: The problem of media list length is not as great as some fear, as EML has already substantially reduced the order effects

The Extended Media List (EML) technique has been highly effective in allowing us to measure a long list of titles with relatively small order effects. Consequently we are limited as to the further improvements that can be made. If this is so, it follows that it is not only how many titles that are shown that matters, but also how they are shown.

Prior to the introduction of EML, the order effects on the British NRS were a matter of some concern, as documented by Ted Whitley in New Orleans. When EML was developed in the early eighties it was with the intention of addressing these concerns, and at the same time enabling the media list to expand.

The introduction of EML in 1984 allowed the media list to expand threefold from around 100 to 300 titles. This was achieved by presenting the titles initially in groups of six, as explained earlier in this paper.

EML substantially reduced the order effects observed. For instance, titles in the group of general monthlies showed an order effect of 38% higher results when shown early in the interview compared with when shown late. With EML this effect was reduced to around 5%.

At the same time there was an increase in the readership estimates recorded. As there were a number of changes to the Survey at the same time, it is not possible to isolate the introduction of EML as the only possible cause of these increases, but given EML's impact upon the order effects we would certainly expect it to be a contributory factor. Gross AIR for titles common to both the 1983 and 1984 surveys showed an overall increase for dailies of 2% and for other titles of 9%.

It is possible to extrapolate from the order effects a theoretical calculation of what might happen if the media list was either reduced or extended. Pym Cornish originally carried out this work in 1990. Prior to commencing the PML development in 1999 we repeated his calculations, which suggested that a reduction of 15 magazine EML cards (as they were at that time) would yield an increase in AIR for the remaining magazines of the 3% magnitude. For newspapers the expected effect would be less than 0.5% magnitude.

Although these calculations will obviously depend on the order effects observed, which do vary from year to year, the point is that the scope for the order effects to be reduced (and the readership estimates to be increased) is limited.

Hypothesis 2: PML has had little effect because it removes titles that are irrelevant to the respondent

Prior to commencing the PML test, we speculated that an improvement in relevance to the respondent might bring about an improvement to the quality of the data obtained. As PML reduces the reams of irrelevant titles the respondent is asked to work through, this in turn might reduce the risk of the respondent being distracted or irritated. However, as we have seen, there is no evidence that improved relevance has brought about any discernible change to the data obtained.

Turning this on its head, we now speculate whether it is precisely because PML removes mainly irrelevant titles that there is no discernible change in the data. With the EML technique it is relatively quick and easy for respondents to dismiss those EML screens containing titles that are completely irrelevant. Therefore there is relatively little 'improvement' in not showing these irrelevant screens in the first place.

One might hypothesise that the respondents least likely to make an erroneous claim (in the negative or positive) are those who are either regular readers, or for whom the titles are completely irrelevant. Conversely those more likely to make erroneous claims are those who have read something in a particular genre of publications, but who are not regular or involved readers. By and large PML will continue to question these irregular/occasional readers about the genres concerned.

Bear in mind that PML technique is to attempt to predict not just who will be a regular reader, but anyone who might have read a title in a particular cluster at some point in the past year. Consequently the strata definitions are relatively broad.

In other words, most irregular and occasional readers will still be shown clusters where there is a possibility they might have read a title at some point. The fact that they have seen fewer groups of completely irrelevant titles may not necessarily improve their ability and interest in identifying and claiming particular titles within genres in which they have a marginal interest.

Hypothesis 3: The length and load of the interview as a whole is as/more important than the number of titles shown

So far we have focused on Title Load per Respondent, i.e. the number of titles asked about. However, the load on the respondent is also comprised of the 'rest of the interview', including the batteries of questions asked about the titles they screen in and all the non-readership questions included to obtain demographic or marketing/lifestyle information. The interview as an entirety, and the time it takes to be administered, is very much relevant.

When considering the total interview load, it is worth bearing in mind the interviewer as well as the respondent. Unlike the respondent, the interviewer is well aware of the extent of the interview that must be completed once the initial media list has been screened. This knowledge will to some extent influence how they conduct the interview (e.g. trying to get a respondent who is short of time through the questions as fast as possible). It may also affect their enthusiasm for the task if they perceive the interview as repetitious and boring.

The potential dangers of interviewers trying to compress the interview to an 'acceptable' time, and what negative impact this might have upon the readership estimates have been discussed at previous Symposia, as, for instance, in Wim van der Noort and Costa Tchaoussoglou's paper 'The Importance of Being Constant: the effects of questionnaire overload' in Berlin.

There is ample evidence of interviewers fitting interviews to the available time. We believe that ag.ma's experiment to cut their media list by a third has resulted in very little time saving to the interview. Instead, interviewers spend more time over a smaller number of readership questions. In the UK we have seen our interview time fall from 36 to 26 minutes over the past ten years, and this may in part reflect pressure to 'fit' the interview to the diminishing amount of time respondents are willing to give.

As the PML technique involves the majority (90%+) of positive readership claims being collected directly it was never expected to have much impact on the interview length, and, as we have seen, only a minute or so of interview time has been saved. If there is pressure on interviewers to fit questions to the time available, little has been done to relieve that pressure (or the repetitious nature of the interview).

Conclusions

The test showed that the PML methodology is successful in reducing TLR. It also passed the evaluation criteria set by clients (which were essentially that the PML sample should not record estimates significantly lower than the control sample).

There was no evidence that the PML methodology changed the estimates obtained by reducing the media list load. There is also no clear evidence of any other improvements to the quality of the data or the interview experience more generally.

There is a negative (but entirely expected) finding to take into account, namely the loss of some precision. It was inevitable that the confidence limits attached to the estimates would increase (primarily because fewer respondents were asked about each title, but also as a result of the imputation process).

Given that a loss in precision must be set against no obvious improvements in the data quality, it seems unlikely that PML will be adopted, however a final decision has yet to be made at the time of writing.

All three of the hypotheses outlined above are potentially relevant in explaining why the application of PML has not made any clear difference to the data obtained. In particular, we believe EML is a powerful tool in reducing order effects and limits the scope for further substantial improvements.

In many ways this is a reassuring conclusion.

We are wary, however, of the inference that as reducing the media list hasn't made much difference, so extending with the addition of more titles or questions than presently measured wouldn't make much difference either. If the list was extended naturally we expect order effects to increase to some extent. Here the concern is, as we have seen, not only the overall order effects but also that fact that some titles will be more vulnerable than others.

It is important to consider, however, not just the actual number of titles shown, but also the interview load as a whole. A key aspect of this is how long it takes to administer the interview, bearing in mind that the tolerance for interview length will not be unconnected to how enjoyable/worthwhile the interview is perceived to be. If interviewers are fitting the interview content, as far as possible, to the time they perceive to be available, then making further demands upon on that time may have an exponentially deleterious effect.

Instead the challenge for the future is to think about how we streamline our interviews, both in order to adapt to new methodologies but also, crucially, to make the best possible use of such time as respondents are prepared to spare us. It may well be that PML has a role to play in this process, particularly in respect of electronic self-completion methodologies (such as online) where there is no interviewer to take the respondent quickly through the initial media list.

References

Whitley, Edward. (1981) Some Rotation Effects in the British Survey. Worldwide Readership Symposium, New Orleans

Cornish, Pym. (1990) The Length of the NRS Media List. Evidence on order effects and Recommendations for January 1991 Title List. Prepared for JICNARS (Not in the public domain)

Van der Noort, Wim and Tchaoussoglou, Costa. (1995) The Importance of Being Constant: the effects of questionnaire overload. Worldwide Readership Symposium, Berlin

Page, Katherine. (2001) Personalised Media Lists. Worldwide Readership Symposium, Venice

Brown, Michael. (2005) The Personalised Media List (PML) Test: Some further comments on the Apr-Dec 2004 data. Prepared for NRS Ltd (Not in the public domain)